

STIC-Biotech/ChemLib

From: Whiteman, Brian  
Sent: Friday, January 27, 2006 8:28 AM  
To: STIC-Biotech/ChemLib  
Subject: seq search

09/889,874

SEQ ID NO: 22 and 23

1) search us patents and published us patent application databases

Thank you,

Brian Whiteman  
Remsen, 2D14  
mail box 2C18  
Patent Examiner - Art Unit 1635  
United States Patent and Trademark Office  
(571) 272-0764

\*\*\*\*\*  
Point of Contact:  
Alexandra Wacławiw  
Technical Info. Specialist  
211 6A02 Tel 308-441  
Searcher: \_\_\_\_\_  
Searcher Phone: \_\_\_\_\_  
Date Searcher Picked up: 1-30  
Date completed: 6-1-30  
Searcher Prep Time: \_\_\_\_\_  
Online Time: 8

\*\*\*\*\*  
Type of Search  
NA# \_\_\_\_\_ AA# 2  
S/L: \_\_\_\_\_ Oligomer: \_\_\_\_\_  
Encode/Transl: \_\_\_\_\_  
Structure #: \_\_\_\_\_ Text: \_\_\_\_\_  
Inventor: \_\_\_\_\_ Litigation: \_\_\_\_\_

\*\*\*\*\*  
Vendors and cost where applicable  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
QUESTEL/ORBIT: \_\_\_\_\_  
LEXIS/NEXIS: \_\_\_\_\_  
SEQUENCE SYSTEM: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (Specify): \_\_\_\_\_

THIS PAGE LEFT BLANK

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: January 30, 2006, 09:40:03 : Search time 25.6283 seconds  
(without alignments)  
5726.190 Million cell updates/sec

Title: US-09-889-874A-22

Perfect score: 1784  
Sequence: 1 FTLRBDSMDWGVSTFNV.....ILFHAGVNSDPYLSSTIVY 334

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : A\_Geneseq\_21.\*  
1: geneseq219808:\*  
2: geneseq219809:\*  
3: geneseq220008:\*  
4: geneseq220018:\*  
5: geneseq220028:\*  
6: geneseq220038:\*  
7: geneseq220038:\*  
8: geneseq220048:\*  
9: geneseq220058:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1784	100.0	334	3	AAV95706
2	145.5	8.2	341	6	ABM67282
3	126	7.1	892	6	ABU42557
4	126	7.1	892	8	AD084849
5	126	7.1	892	8	AD084849
6	126	7.1	930	5	ABP40469
7	126	7.1	930	5	ABP40469
8	126	7.1	991	3	AAV83171
9	126	7.1	991	3	AAV70120
10	124	7.0	1092	7	ABM79019
11	119	6.7	893	7	ADV16755
12	119	6.7	893	7	ADV16755
13	119	6.7	899	7	ADH68812
14	118.5	6.6	670	6	ABU42520
15	118.5	6.6	773	9	ABU42520
16	115.5	6.5	855	5	ABR88573
17	115.5	6.5	1411	6	ADA09358
18	115.5	6.5	1411	6	ADA09358
19	115.5	6.5	2835	5	ABR88574
20	115.5	6.5	2835	5	ABR88574
21	114	6.4	360	8	ABR55594
22	114	6.4	360	9	ABR55594
23	114	6.4	376	8	ADK48623
24	112.5	6.3	1436	6	ADA09354

25	112.5	6.3	1436	8	ADR46771	Adi46771 H. influe
26	111	6.2	369	6	ABU01334	Abu01334 S. pneumo
27	109	6.1	807	3	ABR18311	ABR18311 Plasmodiu
28	109	6.1	807	7	ABO23607	ABO23607 Plasmodiu
29	108.5	6.1	753	8	ADU07705	Adu07705 Amino aci
30	108	6.1	626	3	AAQ30820	AAQ30820 Arabidops
31	108	6.1	626	6	ADT56584	ADT56584 Plant pol
32	108	6.1	978	6	ABU74332	Abu74332 Protein e
33	107.5	6.0	324	2	AAV19983	AAV19983 B. burgdo
34	107.5	6.0	343	2	AAV19982	AAV19982 B. burgdo
35	107.5	6.0	753	8	ADX91694	Adx91694 Plant ful
36	107.5	6.0	1038	5	ABU15903	Abu15903 Protein e
37	107	6.0	1072	5	ABR54963	ABR54963 Lactococc
38	107	6.0	1114	2	AAV21999	AAV21999 M17 anti
39	107	6.0	1449	6	ABU37787	Abu37787 Protein e
40	107	6.0	1449	9	ABR1527	ABR1527 Microbial
41	106.5	6.0	719	8	ADK93594	Adk93594 Plant ful
42	106	5.9	258	2	AAV33726	AAV33726 Photorhab
43	106	5.9	1430	5	AAU98042	AAU98042 S. mutans
44	105.5	5.9	991	8	ADU25478	Adu25478 L. acidop
45	105	5.9	507	3	ABR18173	ABR18173 Plasmodiu

## ALIGNMENTS

RESULT 1	AAV95706	AAV95706 standard; protein; 334 AA.
ID	AAV95706	AAV95706 standard; protein; 334 AA.
AC	AAV95706	AAV95706
XX	AAV95706	AAV95706
DT	25-OCT-2000	(first entry)
XX	25-OCT-2000	(first entry)
DE	Cosmid CHIRMS encoded protein p13-1f.	
XX	Cosmid CHIRMS; nematocyst; nematode; biological control agent;	
KW	transgenic plant; helminthiasis; p13-1f.	
OS	Xenorhabdus bovienii.	
XX	Xenorhabdus bovienii.	
PN	W0200042855-A1.	
XX	W0200042855-A1.	
PD	27-JUL-2000.	
XX	27-JUL-2000.	
PF	24-JAN-2000; 2000MO-GB000219.	
XX	24-JAN-2000; 2000MO-GB000219.	
PR	22-JAN-1999; 99GB-00001499.	
XX	22-JAN-1999; 99GB-00001499.	
PA	(HORT-) HORTICULTURE RES INT.	
XX	(HORT-) HORTICULTURE RES INT.	
PT	Morgan JAW, Jarrett P, Ellis D, Ousley MA;	
XX	Morgan JAW, Jarrett P, Ellis D, Ousley MA;	
XX	WPI; 2000-499157/44.	
DR	N-PSDB; AAAS0029.	
XX	N-PSDB; AAAS0029.	
PT	Novel composition used to control parasitic nematodes, especially in a	
XX	Novel composition used to control parasitic nematodes, especially in a	
PT	plants such as maize, cotton, soy, and rice, comprises a bacterium which	
XX	plants such as maize, cotton, soy, and rice, comprises a bacterium which	
PS	is a symbiont of an entomopathogenic nematode.	
XX	is a symbiont of an entomopathogenic nematode.	
XX	Example 6; Page 42; 74pp; English.	
CC	The present sequence is that of protein p13-1f encoded by an open reading	
CC	frame identified in cosmid CHIRMS (see AAAS0029). CHIRMS was obtained by	
CC	ligating Xenorhabdus bovienii strain 173 (NCIMB 40986) Sau3A-digested DNA	
CC	fragments into the BamHI site of the Stratagene cosmid vector Supercoi,	
CC	packaging into Escherichia coli XL Blue 1, and screening for nematocidal	
CC	activity against Caenorhabditis elegans. Analysis of the DNA indicated a	
CC	number of open reading frames for which the corresponding protein	
CC	sequences were determined (see AAV95685-95735). Nematodes can be	
CC	controlled through the use of bacteria associated symbiotically with an	
CC	entomopathogenic nematode. Such bacteria include Xenorhabdus and	
CC	Photorhabdus spp. such as X. bovienii strain 173. The symbiont bacteria,	

CC an engineered bacterium, or a nematocidal protein obtained from such  
CC bacteria, particularly P13-1f or P14-2f (see AAY95707) can be used to  
CC control helminthiasis in a human or domesticated animal or for the  
CC control of plant pathogen nematodes. Also claimed are vectors for  
CC expressing nematocidal proteins in host cells, and transgenic plants  
XX

Sequence 334 AA;

Query Match 100.0%; Score 1784; DB 3; Length 334;  
Best Local Similarity 100.0%; Pred. No. 5.7e-150; Mismatches 0; Gaps 0;  
Matches 334; Conservative 0; Indels 0;

CC 1 FTLRSDMSDWTGVTNFVILETGLDNCNIYANGNMGVINITPTDDEGNFVDIDVT 60  
CC |||||  
CC 1 FTLRSDMSDWTGVTNFVILETGLDNCNIYANGNMGVINITPTDDEGNFVDIDVT 60  
CC |||||  
CC 61 LNDNIKIVDYIDGSDIDGSDGMPYTGPNENYITPNSQSYLKSNSQITQIKRYVSCS 120  
CC |||||  
CC 61 LNDNIKIVDYIDGSDIDGSDGMPYTGPNENYITPNSQSYLKSNSQITQIKRYVSCS 120  
CC |||||  
CC 121 NTSRLRTKSPSAKVTTSKGKVISITONSINSRVVINAIDATNFTDDELRTKRETNQ 180  
CC |||||  
CC 121 NTSRLRTKSPSAKVTTSKGKVISITONSINSRVVINAIDATNFTDDELRTKRETNQ 180  
CC |||||  
CC 181 SYTSKSTNSLYVHTWTIPRSLKIQWRMEDYNNNGWTWASQCYKXGADGSGSESTWLA 240  
CC |||||  
CC 181 SYTSKSTNSLYVHTWTIPRSLKIQWRMEDYNNNGWTWASQCYKXGADGSGSESTWLA 240  
CC |||||  
CC 241 AGSIFPPGNTDGLMDNDIALSGMAHKSYNVDTGINQLSFRRIIGKGSWYVNIISGLDRG 300  
CC |||||  
CC 241 AGSIFPPGNTDGLMDNDIALSGMAHKSYNVDTGINQLSFRRIIGKGSWYVNIISGLDRG 300  
CC |||||  
CC 301 HAVIIIDQYGNKRIILFHAGYENS DPLYSSSIYV 334  
CC |||||  
CC 301 HAVIIIDQYGNKRIILFHAGYENS DPLYSSSIYV 334  
CC |||||

RESULT 2

ABM67282  
ID ABM67282 standard; protein; 341 AA.

XX ABM67282;

DT 20-NOV-2003 (first entry)

XX Photorhabdus luminescens protein sequence #379.

XX Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;  
XX detection; food; gene expression; plant; animal; microorganism; toxin;  
XX antibiotic; biopesticide; virulence factor; disease model; plague;  
XX whooping cough.

OS Photorhabdus luminescens.

XX MO200294867-A2.

XX 28-NOV-2002.

XX 07-FEB-2002; 2002MO-IB003040.

XX 07-FEB-2001; 2001FR-00001659.

XX (INSP ) INST PASTEUR.  
XX (CNRS ) CNRS CENT NAT RECH SCI.

XX Duchaud E, Taourit S, Glaeser P, Frangeul L, Kunst F, Danchin A;  
XX Buchrieser C;

XX WPI; 2003-148459/14.

XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,  
XX useful e.g. as therapeutic antimicrobials and agricultural pesticides.

PS Claim 2; SEQ ID NO 379; 1205bp; French.

XX The invention relates to the isolation of genes and their encoded  
CC proteins from Photorhabdus luminescens. The isolated sequences are  
CC sources of probes and primers for detecting the genome of P. luminescens  
CC and related species; to study polymorphisms; for gene analysis and for  
CC detection/amplification of the genes. Antibodies (Ab) raised against the  
CC polypeptides encoded by the genes are used for detection/identification  
CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that  
CC carry a gene-containing vector are used to select compounds that  
CC modulate, regulate, induce or inhibit expression of the genes in plants,  
CC animals or microorganisms other than P. luminescens and are able to alter  
CC response or sensitivity to toxins and antibiotics produced by P.  
CC recombinant production of the proteins, particularly toxins and  
CC antibacterials useful as insecticides, bactericides and fungicides. The  
CC genes, proteins, vectors containing the genes and Ab are also useful  
CC therapeutically (to treat microbial infection by bacteria or fungi that  
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as  
CC biopesticides. Other uses of the genes and the proteins are as virulence  
CC factors and for identifying targets of human diseases for which P.  
CC luminescens is a model (particularly plague and whooping cough). This  
CC sequence represents one of the isolated P. luminescens proteins  
XX

Sequence 341 AA;

Query Match 8.2%; Score 145.5; DB 6; Length 341;  
Best Local Similarity 22.9%; Pred. No. 0.00041;  
Matches 74; Conservative 59; Mismatches 137; Indels 53; Gaps 15;

CC 29 NIYANGNMGVINITPTD-DEGNFVDIDVTLNDNIK-----IVDYIDGSDIDGSD 80  
CC |||||  
CC 20 NFAVANGHQCMVNTSVLKQEKYKNGDWYKL--ALSDAEKKSIVQVALSDSLIYDQLKMP 76  
CC |||||  
CC 81 GWPRTGNPNY-----NTINSOSYS-----LKSNEGQ-ITQIKRYVSC-SNT 122  
CC |||||  
CC 77 GWTTPDARNKYDGLGLNGVYSADI FTBEPVIRNAGDCTTBNQNSVSEIILCYVSS 136  
CC |||||  
CC 123 SRLRTKSPSAKVT--TTSQVVISITQNSIN-----SRVIVNAIDATNFTDDELRTKRETR 176  
CC |||||  
CC 137 NRTSTBYLMAKMTPEIDNGKRTLTITMNSVGDEVPDSKILKALAPYANNOQHESNITL 196  
CC |||||  
CC 177 FENQSYTSKRSSTN--SLVYHTWTIPRSLKIQWRMEDYNNNGWTWASQCYKXGADGSGSE 234  
CC |||||  
CC 197 FDKTEBPT-KSDTHGQITNLYRWTLFYLRLILBGNDRVNNIYVLGRS-----SS 245  
CC |||||  
CC 235 STRWLAAGSIFPPGNTDGLMDNDIALSGMAHKSYNVDTGINQLSFRRIIGKGSWYVNI 294  
CC |||||  
CC 246 DDRFLTRARVFKGT-SYVARRDMSGGCVWDYSYDVTPDTQLAAEVLHVTGSGSWTTGY 304  
CC |||||  
CC 295 SGLDRGHAVIIIDQYG--NRYRI 315  
CC |||||  
CC 305 --VDGYHDVTIIDNYGCGHFKRI 325  
CC |||||

RESULT 3

ABU42557  
ID ABU42557 standard; protein; 892 AA.

XX ABU42557;

DT 19-JUN-2003 (first entry)

XX Protein encoded by Prokaryotic essential gene #28084.

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX Staphylococcus epidermidis.

XX MO200277183-A2.

XX 03-OCT-2002.



```

XX 21-MAR-2002; 2002ZWO-US009107.
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-00342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
PA (EILIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JW;
PI Wall D, Travick JD, Carr GJ, Yamamoto R, Foreynh RA, Xu HH;
XX WPI: 2003-029926/02.
DR N-PSDB; ACA46427.
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
PS Claim 25; SEQ ID NO 70481; 1766bp; English.
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 623 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 892 AA;
Query Match 7.1%; Score 126; DB 6; Length 892;
Best Local Similarity 21.0%; Pred. No. 0.089;
Matches 78; Conservative 51; Mismatches 141; Indels 102; Gaps 16
QY 9 SDWTGVSFTNVILETGLDNCNIYANGLMNIG--VIINTPTDDEGNFADIDVT----- 60
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 398 SMFTNIDTKNHTVEQ-----TIYINPLRYSAKETWNVISGNDESESTI-IDSTIIKIVK 451
QY : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 452 VGDNONLPDSNRITDYSEIEDT-NDVYQAQLNNDNVINFGNIDSPYIIKIVSKYDEPK 510
QY : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 98 QSYSLKSENSQITQIKRYVSCSNTRSRATKSPSAKV--TTTSG-----KVYSI-- 144
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 511 DDYTTIQQVTVMQTINNEY-----TGEPRITASVDYMTIAFSTSSGGCGQDLPREKTVKIGD 565
QY : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 145 -----TQNSINSSRRVAINAIDATNFDTDELRTTKETRFENQSYTHSKSSTNSL-Y 193

```

[illegible]

CC antigens in a human or animal patients suspected of containing the  
CC antigens or antibodies, in preventing or reducing infection of medical  
CC devices and prostheses caused by such organisms, and in treating or  
CC preventing infections in highly susceptible groups such as premature  
CC newborns, AIDS and debilitated cancer patients, and bone marrow  
CC transplantation. The present sequence is that of a surface anchored LPXTG  
CC protein identified using the method of the invention.

XX  
SQ Sequence 892 AA;

Query Match 7.1%; Score 126; DB 8; Length 892;

Best Local Similarity 21.0%; Pred. No. 0.089;

Matches 78; Conservative 51; Mismatches 141; Indels 102; Gaps 16;

QY 9 SDMTGVSTFNVILETGLDNCNIYANGLMIG--VIINITPTDDEGNFVDIDVT----- 60

DB 398 SMFTNIDTKNHTVEQ-----TIIYNPLRYSAKETNVNISGNGDEGSTI-IDDSIIIRVYK 451

QY 61 -----LNDNIKIYDIYDSDIDSDGDMFYTGPNFENVNT-----IPNS 97

DB 452 VGDNQNLPSNRIRYDSEYEDVT-NDVYAOLGNNNDVNIINGNIDSPYIIKVISKYPNK 510

QY 98 QSYSLKSENSQITQIKRYVSCNTSRLRTKFSFAKV--TTTSG-----KVISI-- 144

DB 511 DDYTTIQYVTVMQTINNEY-----TGERFTASYNNTIAFSTSSGCGGDLPEPKTYKIGD 565

QY 145 -----TQNSINSSRVVINAIDATNFTDDELRTTKETRFENQSYTHKSSTNSL-Y 193

DB 566 YWEDVDKDGIQNTNDEKPLSNVLVLTTPDG--TSKSVRTDEBKYPDGLKNGILTY 622

QY 194 VHTWTIPRSLKLGQWRWEDY-----NNGWTMAQSCYYTKGADGSEST 236

DB 623 KITETP-----EGYPTPLKHSQTNPALDSBGSVWVTINQDDMTIDSGFYQTP 672

QY 237 RWLAAGSIFPPGNYDGLMNDIALSGMAKSYVVDGIGNLSFTRIIGKFSWVNIISG 296

DB 673 KYSLGNYVWDTNKGIGQDDEKGISGV--KVTLKDENGNIISTTTDENG--KYQFDN 727

QY 297 LDRGHAVIITDQ 308

DB 728 LNSGNVIHFDPK 739

RESULT 5  
ADS20651 standard; protein; 892 AA.

XX  
AC ADS20651;

XX  
DT 30-DEC-2004 (first entry)

XX  
DE S. epidermidis hyperimmune serum reactive antigen protein - SEQ ID 52.

XX  
KW antigen; antibiotic resistance; antibacterial; vaccine; gene therapy.

XX  
OS Staphylococcus epidermidis.

XX  
PN WO2004087746-A2.

XX  
PD 14-OCT-2004.

XX  
PF 31-MAR-2004; 2004WO-EP003398.

XX  
PR 31-MAR-2003; 2003EP-00450078.

XX  
PA (INTE-) INTERCELL AG.

XX  
PI Meinke A, Min Bui D, Nagy E;

XX  
DR WPI; 2004-729219/71.

XX  
N-PSDB; ADS20620.

XX  
PT New nucleic acid molecules encoding hyperimmune serum reactive antigens

PT from Staphylococcus epidermidis, useful for diagnosing, preventing or  
XX treating S. epidermidis infections.

XX  
PS Claim 11, SEQ ID NO 52; 196pp; English.

XX  
CC The invention relates to a novel isolated nucleic acid molecule encoding  
CC a hyperimmune serum reactive antigen or its fragment. Staphylococci are  
CC commonly associated with human disease. Both Staphylococcus epidermidis  
CC and Staphylococcus aureus have become resistant to many commonly used  
CC antibiotics, most importantly methicillin (MRSA) and vancomycin (VISA).

CC Drug resistance is an increasingly important public health concern and  
CC prevention therapies to combat staphylococci infection must be developed in  
CC preparation for a time when such infections may be untreatable by  
CC antibiotics. The molecules of the invention demonstrate antibacterial  
CC activity and may be useful for manufacturing a medicament, such as a  
CC vaccine, for treating or preventing S. epidermidis infections, possibly  
CC via gene therapy. The antigen or its fragment may also be used for  
CC generating an anticaline peptide binding to the hyperimmune serum  
CC reactive antigen or fragment, for manufacturing a functional nucleic  
CC acid, such as an aptamer or spligamer and for manufacturing a functional  
CC ribonucleic acid, such as a ribozyme, antisense nucleic acid or siRNA  
CC (short interfering RNA). The current sequence is that of a Staphylococcus  
CC epidermidis hyperimmune serum reactive antigen protein of the invention.

XX  
SQ Sequence 892 AA;

Query Match 7.1%; Score 126; DB 8; Length 892;

Best Local Similarity 21.0%; Pred. No. 0.089;

Matches 78; Conservative 51; Mismatches 141; Indels 102; Gaps 16;

QY 9 SDMTGVSTFNVILETGLDNCNIYANGLMIG--VIINITPTDDEGNFVDIDVT----- 60

DB 398 SMFTNIDTKNHTVEQ-----TIIYNPLRYSAKETNVNISGNGDEGSTI-IDDSIIIRVYK 451

QY 61 -----LNDNIKIYDIYDSDIDSDGDMFYTGPNFENVNT-----IPNS 97

DB 452 VGDNQNLPSNRIRYDSEYEDVT-NDVYAOLGNNNDVNIINGNIDSPYIIKVISKYPNK 510

QY 98 QSYSLKSENSQITQIKRYVSCNTSRLRTKFSFAKV--TTTSG-----KVISI-- 144

DB 511 DDYTTIQYVTVMQTINNEY-----TGERFTASYNNTIAFSTSSGCGGDLPEPKTYKIGD 565

QY 145 -----TQNSINSSRVVINAIDATNFTDDELRTTKETRFENQSYTHKSSTNSL-Y 193

DB 566 YWEDVDKDGIQNTNDEKPLSNVLVLTTPDG--TSKSVRTDEBKYPDGLKNGILTY 622

QY 194 VHTWTIPRSLKLGQWRWEDY-----NNGWTMAQSCYYTKGADGSEST 236

DB 623 KITETP-----EGYPTPLKHSQTNPALDSBGSVWVTINQDDMTIDSGFYQTP 672

QY 237 RWLAAGSIFPPGNYDGLMNDIALSGMAKSYVVDGIGNLSFTRIIGKFSWVNIISG 296

DB 673 KYSLGNYVWDTNKGIGQDDEKGISGV--KVTLKDENGNIISTTTDENG--KYQFDN 727

QY 297 LDRGHAVIITDQ 308

DB 728 LNSGNVIHFDPK 739

RESULT 6  
ABP40469 standard; protein; 930 AA.

XX  
ID ABP40469

XX  
AC ABP40469;

XX  
DT 24-JUL-2002 (first entry)

XX  
DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5314.

XX  
KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;

XX  
KM antibacterial; gene therapy.

XX  
OS Staphylococcus epidermidis.

XX  
PT

XX US6380370-B1.  
 PN 30-APR-2002.  
 PD 13-AUG-1998; 98US-00134001.  
 PF 14-AUG-1997; 97US-0055779P.  
 PR 08-NOV-1997; 97US-0064964P.  
 XX (GENO-) GENOME THERAPEUTICS CORP.  
 PA Doucette-Stamm LA, Bush D;  
 PI WPI: 2002-381255/41.  
 DR N-PSDB; ABN93014.  
 XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis  
 PT polypeptide, useful for diagnosing and treating bacterial infections.  
 XX Disclosure; SEQ ID NO 5314; 267pp; English.  
 PS  
 XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
 CC given in ABP35124 to ABP37960. The S. epidermidis sequences have  
 CC antibacterial activity and can be used in gene therapy. The sequences can  
 CC also be used in the diagnosis and treatment of bacterial infections,  
 CC particularly S. epidermidis infections. The sequences can be used to  
 CC screen for compounds able to interfere with the S. epidermidis life cycle  
 CC or inhibit S. epidermidis infection. N.B. The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from the USPTO web site  
 XX  
 SQ Sequence 930 AA;

Query Match 7.1%; Score 126; DB 5; Length 930;  
 Best Local Similarity 21.0%; Pred. No. 0.095;  
 Matches 78; Conservative 51; Mismatches 141; Indels 102; Gaps 16;  
 QY 9 SDMTGVTSTVAVILETGLDNCNIYANGLMIG--VIINITPTDDEGNFVIDDVT----- 60  
 DB 436 SMFTNIDTKNHTVEQ-----TIYINPLRYSAKETNNVNSGNGEGSTI-IDDSTIIKVKYK 489  
 QY 61 -----LNDNIKIVDYIDSGSDGWFYTGPNENYNT-----IPNS 97  
 DB 490 VGNQNLPSNRKYDSEYEDVT-NDYVQLGNNNDVNNFGNIDSPYIIKVISKIDPNK 548  
 QY 98 QSYSLKSENSQITQIKRYVSCSWTSRLRTKSFSAKV--TTTSG-----KVISI-- 144  
 DB 549 DDTTITQIYTMGTINNEY-----TGFRTASVYNTIATFSTSSGCGGGLPPEKTYKIGD 603  
 QY 145 -----TQNSINSSRVAINAIDATNTDELRTTKETREPENOSYTHSKSTNSL-Y 193  
 DB 604 YWEDVDKQDIGNTNDKEKELSNVLVLTYPDG---TSKSVRTDBEKGYPFDGLKNGLTLY 660  
 QY 194 VHTWTIPRSKLQNKWMEY-----NNGWTAQSCYVYTKGADGSEST 236  
 DB 661 KTFETP-----EGYTPILKHSCTNPALDSEGNASVWVITGDDMTTDSGFQTP 710  
 QY 237 RMLAAGSIFPPGVNDGLNDIALSGMAHSYVNDVIGIQLSFTRIIGKGFVYVNI 296  
 DB 711 KYELGNVYVWDTNKDGIQSGDKEKISGV--KVTIKDENGMIIITTTDENG--KIQFDN 765  
 QY 297 LDRGHAVIITDQ 308  
 DB 766 LNSGNVYVHDK 777

RESULT 7  
 ADS06014  
 ID ADS06014 standard; protein; 930 AA.  
 AC ADS06014;

XX 04-NOV-2004 (first entry)  
 DT Staphylococcus epidermis polypeptide seqid 5309.  
 DE antibacterial; vaccine; antisense therapy; Staphylococcus epidermidis;  
 XX recombinant expression vector; infection; computer readable medium;  
 KM computer based system.  
 XX Staphylococcus epidermidis:  
 OS US2004147734-A1.  
 PN 29-JUL-2004.  
 PD 01-DEC-2003; 2003US-00724972.  
 PF 08-NOV-1997; 97US-0064964P.  
 PR 13-AUG-1998; 98US-00134001.  
 PR 29-NOV-1999; 99US-00450969.  
 XX (DOUC/) DOUCETTE-STAMM L.  
 PA (BUSH/) BUSH D.  
 PI Doucette-Stamm L, Bush D;  
 DR WPI: 2004-580138/56.  
 DR N-PSDB; ADS02242.  
 XX  
 XX Claim 17; SEQ ID NO 5309; 741pp; English.  
 PT The invention describes an isolated nucleic acid comprising a nucleotide  
 PT sequence with any of 3772 fully defined nucleotide sequences (SEQ ID NO:  
 CC 1-3772) and encoding an Staphylococcus epidermidis polypeptide with any  
 CC of 3772 fully defined amino acid sequences (SEQ ID NO: 3772-7544) as  
 CC given in the specification. Also described are: a recombinant expression  
 CC vector; a cell comprising a recombinant expression vector of (1);  
 CC producing an S. epidermidis polypeptide; an isolated nucleic acid  
 CC comprising a nucleotide sequence of at least 8 nucleotides in length; a  
 CC vaccine composition for prevention or treatment of an S. epidermidis  
 CC infection, comprising a nucleic acid cited above and a carrier; treating  
 CC a subject for S. epidermidis infection; a recombinant or substantially  
 CC pure preparation of an S. epidermidis polypeptide or its fragment; a  
 CC vaccine composition for prevention or treatment of an S. epidermidis  
 CC infection; detecting the presence of a Staphylococcus nucleic acid in a  
 CC sample; a computer readable medium having recorded in it the nucleotide  
 CC sequences with SEQ ID NO: 1-3772 or its fragments; a computer based  
 CC system for identifying fragments of the Staphylococcus genome of  
 CC commercial importance; a computer based system for identifying fragments  
 CC of the Staphylococcus plasmids of commercial importance; identifying  
 CC commercially important nucleic acid fragments of the Staphylococcus  
 CC genome and/or plasmids; and identifying an expression modulating fragment  
 CC of the Staphylococcus genome and/or plasmids. The methods and  
 CC compositions of the present invention are useful for the diagnosis,  
 CC prevention and/or treatment of an Staphylococcus epidermidis bacterial  
 CC infection. This is the amino acid sequence of a S. epidermis protein of  
 CC the invention.  
 XX  
 SQ Sequence 930 AA;

Query Match 7.1%; Score 126; DB 8; Length 930;  
 Best Local Similarity 21.0%; Pred. No. 0.095;  
 Matches 78; Conservative 51; Mismatches 141; Indels 102; Gaps 16;  
 QY 9 SDMTGVTSTVAVILETGLDNCNIYANGLMIG--VIINITPTDDEGNFVIDDVT----- 60  
 DB 436 SMFTNIDTKNHTVEQ-----TIYINPLRYSAKETNNVNSGNGEGSTI-IDDSTIIKVKYK 489  
 QY 61 -----LNDNIKIVDYIDSGSDGWFYTGPNENYNT-----IPNS 97

```

Db      490 VGDNQNLPSDNRIYDSEYEDVT-NDDYAQLGNNNDVNIINGNIDSPYIIKVISKYPNK 548
Qy      98 QSVSLKSENSQITQIKRYVSCNTSLRTRKSPSAKY--TTTSG-----KVISI-- 144
Db      549 DDYTTIOQTVMQTTINEY-----TGEFRITASYDNTIAFTSSQGGGDLPEPETYKIGD 603
Qy      145 -----TONSSSRVYINADITNFTDDELRTTKETRFENQSYTHSKSTNSL-Y 193
Db      604 YWEDVDVKDGIQTNDNEKPLSNVLVLTYPDG---TSKSVRTDEBKQYDGLKNGLTY 660
Qy      194 VHTWTIRSLKLQWRMEDY-----NNGWTMAQSCYYTKGADGSEST 236
Db      661 KITFETP-----EGYTPTLKHSGTNPALDSEGNVWVTINGODDMTIDSGFYQTP 710
Qy      237 RMLAAGSIFPPGNDGLMNDIALSGMAHKSIVNDGIGQLSPTRIIKGFSWVYINISG 296
Db      711 KYSIGNYWYDYNKDGIGDDEKGISGV--KVTLKDENGNIISTTTDENG--KYQPDN 765
Qy      297 LDRGHAVIIDIQ 308
Db      766 LNSGNVIVHFDK 777

```

RESULT 8  
ID AAV83171 standard; protein; 991 AA.

AAV83171;

24-JUN-2000 (first entry)

Cell wall protein Sdrg.

Sdrg; Sdrg; Sdrg; coagulase negative; staphylococcus; septicemia;  
osteomyelitis; endocarditis; immune response; vaccine; graft; stent;  
intravenous catheter; heart valve; cardiac.

Staphylococcus sp.

Key Location/Qualifiers

Misc-difference 14 /note= "Position encoded by TAG stop codon"

Misc-difference 33 /note= "Position encoded by TGA stop codon"

Misc-difference 964 /note= "Position encoded by TAA stop codon"

Misc-difference 980 /note= "Position encoded by TAG stop codon"

Misc-difference 989 /note= "Position encoded by TAA stop codon"

MO200012689-A1.

09-MAR-2000.

31-AUG-1999; 99WO-US019728.

31-AUG-1998; 98US-0098443P.

25-JAN-1999; 99US-0117119P.

(QUEE-) QUEEN ELIZABETH COLLEGE DUBLIN.

(TEXA) UNIV TEXAS A & M SYSTEM.

WPI, 2000-256637/22.

N-PSDB; AA293534.

Recombinant or synthetic proteins from coagulase-negative staphylococci  
useful for prevention, treatment and diagnosis of staphylococcal  
infections bind soluble and immobilized fibrinogen.

PS Claim 8; Fig 3; 104pp; English.  
XX Isolated Staphylococcus Sdr cell wall proteins which bind both soluble  
CC and immobilized fibrinogen are useful for treating or preventing  
CC coagulase-negative staphylococcal infection such as septicemia,  
CC osteomyelitis or endocarditis, and for inducing immune responses in  
CC patients. The cell wall proteins are also useful for reducing coagulase-  
CC negative staphylococcal infection of indwelling medical devices such as  
CC vascular grafts, vascular stents, intravenous catheters, artificial heart  
CC valves and cardiac assist devices. The cell wall associated proteins are  
CC able to inhibit staphylococcal adhesion to immobilised extracellular  
CC matrix or host cells present on the surface of implanted biomaterials  
XX

Sequence 991 AA:

Query Match 7.1%; Score 126; DB 3; Length 991;

Best Local Similarity 21.0%; Pred. No. 0.1; Indels 102; Gaps 16;

Matches 78; Conservative 51; Mismatches 141; Indels 102; Gaps 16;

9 SDWTGVSTFNVILETGLDNCNIYANGIMIG--VIINIFPTDEGNFVDIDVT----- 60

469 SMFTNIDTKHMYEQ-----TIYNPLRYSAKETNNVNSGNGEGSTI-IDSTIIIVYK 522

61 -----LNDNIKIVDYIDSGSDIDSGEYTGPNENYNT-----IPNS 97

523 VGDNQNLPSDNRIYDSEYEDVT-NDDYAQLGNNNDVNIINGNIDSPYIIKVISKYPNK 581

98 QSVSLKSENSQITQIKRYVSCNTSLRTRKSPSAKY--TTTSG-----KVISI-- 144

582 DDYTTIOQTVMQTTINEY-----TGEFRITASYDNTIAFTSSQGGGDLPEPETYKIGD 636

145 -----TONSSSRVYINADITNFTDDELRTTKETRFENQSYTHSKSTNSL-Y 193

637 YWEDVDVKDGIQTNDNEKPLSNVLVLTYPDG---TSKSVRTDEBKQYDGLKNGLTY 693

194 VHTWTIRSLKLQWRMEDY-----NNGWTMAQSCYYTKGADGSEST 236

694 KITFETP-----EGYTPTLKHSGTNPALDSEGNVWVTINGODDMTIDSGFYQTP 743

237 RMLAAGSIFPPGNDGLMNDIALSGMAHKSIVNDGIGQLSPTRIIKGFSWVYINISG 296

744 KYSIGNYWYDYNKDGIGDDEKGISGV--KVTLKDENGNIISTTTDENG--KYQPDN 798

297 LDRGHAVIIDIQ 308

799 LNSGNVIVHFDK 810

RESULT 9  
ID AAV70120 standard; protein; 991 AA.

AAV70120;

06-JUN-2000 (first entry)

Staph. epidermidis serine-aspartate repeat region protein Sdrg.

Multi-component vaccine; immunostimulatory; antibacterial; MSCRAMM;  
microbial surface components recognising adhesive matrix molecules;  
collagen binding protein; CBP; CNA; fibrinogen binding protein;  
clumping factor A; ClfA; Clumping factor B; ClfB; Fbp;  
fibrinectin binding protein; Staphylococcus infection;  
serine-aspartate repeat region protein; SDR protein; Sdrg.

Staphylococcus epidermidis.

Key Location/Qualifiers

Misc-difference 14 /note= "Encoded by in-frame stop codon TAG"

Misc-difference 33 /note= "Encoded by in-frame stop codon TGA"

Misc-difference 964

```

FT /note= "Encoded by in-frame stop codon TAA"
FT Misc-difference 980
FT /note= "Encoded by in-frame stop codon TAG"
FT Misc-difference 989
FT /note= "Encoded by in-frame stop codon TAA"
XX
XX MO200012131-A1.
XX
XX 09-MAR-2000.
XX
XX 31-AUG-1999; 99WC-US019727.
XX
XX 31-AUG-1998; 98US-0098439P.
XX
XX (INHIT-) INHIBITEX INC.
XX (TEXA-) UNIV TEXAS A & M SYSTEM.
XX (QUEE-) QUEEN ELIZABETH COLLEGE DUBLIN.
XX
XX Patti JM, Foster TJ, Hook M;
XX
XX WPI; 2000-237781/20.
XX N-PSDB; AA251202.
XX
XX Composition used for generating immune response or for inhibiting
XX microbial colonization in an animal comprises antibodies that bind
XX collagen binding protein, fibrinogen binding protein and, optionally,
XX fibrinectin binding protein.
XX
XX Claim 8; Fig 4; 115BP; English.
XX
XX The patent discloses multicomponent vaccines containing selected
XX combinations of bacterial binding proteins termed MCRAMM (microbial
XX surface components recognising adhesive matrix molecules) or their
XX antibodies. A vaccine composition is provided that includes collagen
XX binding protein or peptide, e.g. CNA, a fibrinogen binding protein
XX preferably Clumping factor A (ClfA) or Clumping factor B (ClfB), and
XX optionally a fibrinectin binding protein e.g. FnBP-A. The vaccines are
XX useful for imparting protection against a broad spectrum of
XX Staphylococcal strains and for inhibiting microbial colonization,
XX especially of Staphylococcus aureus, in an animal. The combinations can
XX also be used to select donor blood pools for the preparation of purified
XX blood products for passive immunisation. The present sequence is a serine
XX -aspartate repeat region protein, Sdrg from Staphylococcus epidermidis.
XX The Sdr protein is useful in vaccine preparation in combination with
XX specific bacterial binding proteins. These vaccines can be used to treat
XX a broad spectrum of bacterial infections, including those arising from
XX both coagulase-positive and coagulase-negative bacteria
XX
XX Sequence 991 AA:
XX
XX Query Match 7.1%; Score 126; DB 3; Length 991;
XX Best Local Similarity 21.0%; Pred. No. 0.1;
XX Matches 78; Conservative 51; Mismatches 141; Indels 102; Gaps 16;
XX
XX 9 SDWTGTVFVILETGLDNCNIYANGLMIG--VIINITEPTDDEGNFVDIDVT-----60
XX |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
XX SMTETNIDTKHMYEQ-----TIINPLRYSAKETNVIISNGDEGSTI-IDDTIIKIVK 522
XX
XX 61 -----LNDNIKIVDYIDGSDIGSGDFYTGPNFNEVNT-----IPNS 97
XX |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
XX VGDNQMLPDSNRILYDSEVEDVT-NDDYAQLGNNNDVINFGNIDSPYIIKIVISKYDPNK 581
XX
XX 98 QSVSLKSEMSQTTQIKRYVSCNTSRKTKSFSAKV--TTTGG-----KVSI--144
XX |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
XX DDVTITQCTVWQTTINEY-----TGEFRYASVDNTIAFSTSSGGQGGDLPPERTYKIGD 636
XX
XX 145 -----TQNSINSRVIVNAIDATNFTDELRTTKETRENOGYTSKKSSTNSL-Y 193
XX |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
XX YWVEDVDKQIQNTNDEKPLSNVLVLYTPDG---TSKSVKRDDEBKYPDGLKNGLTY 693
XX
XX 194 VHTWTTPRSKLQKNRWEDY-----NNGWTVAQSCYYRTGAGDGESEST 236
XX |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
XX KITEFP-----EGYTPTLKISGTNPALDSEGNVWVTINGDDWTIDSGFYQTP 743
XX

```

```

XX 237 RMLAGSIPEPGNYDGLMDNDIALSGMAHKSYNVDITGNGLSFTRIIGKFSWYNISG 296
XX |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
XX 744 KYSLGNVWYDYNKKD1QGDDEKISGV--KVTLKDENGNIISTTTDENG---KYQFDN 798
XX
XX 297 LDRGHAVIIDIQ 308
XX |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
XX 799 LNSGNYIVHFDK 810
XX
XX
XX RESULT 10
XX AA#41602
XX ID AA#41602 standard; protein; 1092 AA.
XX
XX AA#41602;
XX
XX 17-OCT-2003 (revised)
XX 22-JUN-1998 (first entry)
XX
XX Staphylococcus epidermidis fibrinogen binding protein FIG.
XX
XX Fibrinogen binding protein; FIG; aggregation; infection;
XX coagulase-negative Staphylococcus; therapy; diagnosis; immunisation;
XX immunogen; vaccine.
XX
XX Staphylococcus epidermidis; strain HB.
XX
XX Key Location/Qualifiers
XX Peptide 1..51
XX /label= Sig_peptide
XX Protein 52..1092
XX /label= Mat_protein
XX Region 52..824
XX /note= "non-repetitive region, harbours fibrinogen
XX binding activity"
XX Region 825..1040
XX /note= "Aap-Ser dipeptide repeat region"
XX Region 1053..1057
XX /note= "cell wall anchoring motif"
XX
XX MO9748727-A1.
XX
XX 24-DEC-1997.
XX
XX 18-JUN-1997; 97WC-SR001091.
XX
XX 20-JUN-1996; 96SE-00002496.
XX
XX (GUSG/) GUSG B.
XX (NITS/) NILSSON M.
XX (FRYK/) FRYKBERG L.
XX (FLOC/) FLOCK J.
XX (LIND/) LINDBERG M.
XX
XX Guss B, Nilsson M, Frykberg L, Flock J, Lindberg M;
XX WPI; 1998-063079/06.
XX N-PSDB; AA#04279.
XX
XX Fibrinogen-binding protein from coagulase-negative Staphylococcus - used
XX for prevention, treatment and diagnosis of Staphylococcus infection.
XX
XX Example 3; Fig 6; 45BP; English.
XX
XX The protein comprises the fibrinogen binding protein (FIG) of coagulase-
XX negative Staphylococcus epidermidis HB. Its amino acid sequence was
XX deduced from the isolated fig gene (see AA#04279). The closest known
XX analogue of FIG is the clumping factor of S. aureus which also binds
XX fibrinogen and promotes bacterial aggregation in serum. Recombinant FIG
XX polypeptides can be expressed in host cells. They are used as immunogens,
XX particularly in vaccines (which may be expressed in vivo) to protect
XX humans and animals against coagulase-negative Staphylococcus infection.
XX Antibodies raised against FIG can be used for passive immunisation. They
XX

```

CC block the adherence of bacteria) and for diagnosis. (Updated on 17-OCT-  
CC 2003 to standardise OS field)  
XX Sequence 1092 AA;

Query Match 7.0%; Score 124; DB 2; Length 1092;

Best Local Similarity 21.0%; Pred. No. 0.18;  
Matches 78; Conservative 50; Mismatches 142; Indels 102; Gaps 16;

```
QY 9 SDMTGVSTFVILETGDNKNIYANGNMIG--VIINITPTDDEGNFVDIDVT-----60
DB 439 SMFTNIDTKNHYEQ-----TIYINPLRYSAKETNNVNSGNGEGSTI-IDDSTIIKVVYK 492
QY 61 -----LNDNIKIVDYIDSGDIDSDGMFYTGPNENEYNT-----IPNS 97
DB 493 VGDNONLPDSNRIVDYSEYEDVT-NDYVQAGNNNDVINFGNIDSPYIIKVISKYDPNK 551
QY 98 QSYSLKSENSQITQIKRYVSCSNTSRKTSFSKAV--TTTSG-----KVISI--144
DB 552 DDYTTIQQVYTMQTTINEY-----TGEFRTASVDNTIAFSTSGQGQGLPPEKTYKIGD 606
QY 145 -----TONSINSRVVINAIDATNFTDDELRTTKETREFENOSYTHSKSTNSL-Y 193
DB 607 YWEDVDKDGIQNTDNEKPLSNVLTLYTPDG---TSKSVRTDEDEKQYFDGLKNGLTY 663
QY 194 VHTWTIPRSLKIQWREMDY-----NNGWTMAQSCYYTKGADGSGEST 236
DB 664 KITETP-----EGYPTLKHSGTNPALDSEGNVWVTINQDDMTIISGFYQTP 713
QY 237 RMLAAGIFPPGNYDGLMNDIALSGMAHKSYNVDGIGNQLSFTRIIGKFSVWVVISG 296
DB 714 KYSIGNYWYDNTKDGIGDDEKIGISGV--KVTLKDENGNIISTTTDENG---KYQFDN 768
QY 297 LDRGHAVIIDIQ 308
DB 769 LNSGNYIVHFDK 780
```

## RESULT 11

ABM79019  
ID ABM79019 standard; protein; 1092 AA.

XX ABM79019;

DT 15-JAN-2004 (first entry)

XX Staphylococcus epidermidis polypeptide.

XX Infection; antibacterial; vaccine.

XX Staphylococcus epidermidis.

XX WO2003076470-A1.

XX 18-SEP-2003.

XX 05-MAR-2003; 2003WO-US006415.

XX 05-MAR-2002; 2002US-0361324P.

XX (INH1-) INHIBITEX INC.

XX (TEXA) UNIV TEXAS A & M SYSTEM.

XX Patti JM, Hutchins JT, Hall A, Domanski P, Patel P, Hook M,  
PI Robbins J, Vernachio J, Bowden MG;

XX WPI; 2003-722324/68.

XX New antibody recognizing a Staphylococcus epidermidis protein comprising  
PT Sdrc N1N23, Sdrc N2N3 or SdrcR2 useful for preparing a composition for  
FT treating or preventing a coagulase-negative Staphylococcal infection.

XX Claim 16; Page 36-37; 78pp; English.

XX The present sequence comprises the protein sequence of a polypeptide of a  
CC coagulase-negative Staphylococcus epidermidis. A claimed monoclonal  
CC antibody recognises this protein and is used in a claimed method of  
CC treating or preventing a coagulase-negative staphylococcal infection in a  
CC human or animal, e.g. a nosocomial coagulase-negative staphylococcal  
XX infection in low birth weight infants

XX Sequence 1092 AA;

Query Match 7.0%; Score 124; DB 7; Length 1092;

Best Local Similarity 21.0%; Pred. No. 0.18;  
Matches 78; Conservative 50; Mismatches 142; Indels 102; Gaps 16;

```
QY 9 SDMTGVSTFVILETGDNKNIYANGNMIG--VIINITPTDDEGNFVDIDVT-----60
DB 439 SMFTNIDTKNHYEQ-----TIYINPLRYSAKETNNVNSGNGEGSTI-IDDSTIIKVVYK 492
QY 61 -----LNDNIKIVDYIDSGDIDSDGMFYTGPNENEYNT-----IPNS 97
DB 493 VGDNONLPDSNRIVDYSEYEDVT-NDYVQAGNNNDVINFGNIDSPYIIKVISKYDPNK 551
QY 98 QSYSLKSENSQITQIKRYVSCSNTSRKTSFSKAV--TTTSG-----KVISI--144
DB 493 VGDNONLPDSNRIVDYSEYEDVT-NDYVQAGNNNDVINFGNIDSPYIIKVISKYDPNK 551
QY 98 QSYSLKSENSQITQIKRYVSCSNTSRKTSFSKAV--TTTSG-----KVISI--144
DB 552 DDYTTIQQVYTMQTTINEY-----TGEFRTASVDNTIAFSTSGQGQGLPPEKTYKIGD 606
QY 145 -----TONSINSRVVINAIDATNFTDDELRTTKETREFENOSYTHSKSTNSL-Y 193
DB 607 YWEDVDKDGIQNTDNEKPLSNVLTLYTPDG---TSKSVRTDEDEKQYFDGLKNGLTY 663
QY 194 VHTWTIPRSLKIQWREMDY-----NNGWTMAQSCYYTKGADGSGEST 236
DB 664 KITETP-----EGYPTLKHSGTNPALDSEGNVWVTINQDDMTIISGFYQTP 713
QY 237 RMLAAGIFPPGNYDGLMNDIALSGMAHKSYNVDGIGNQLSFTRIIGKFSVWVVISG 296
DB 714 KYSIGNYWYDNTKDGIGDDEKIGISGV--KVTLKDENGNIISTTTDENG---KYQFDN 768
QY 297 LDRGHAVIIDIQ 308
DB 769 LNSGNYIVHFDK 780
```

## RESULT 12

ADV16755  
ID ADV16755 standard; protein; 893 AA.

XX ADV16755;

DT 24-FEB-2005 (first entry)

XX E. faecalis V583 hyperimmune serum reactive antigen protein - SEQ ID 450.

XX antigen; antibacterial; vaccine; enterococcus infection; infection;  
KM pharyngitis; impetigo; rheumatic fever; antipyretic; antineumatic;  
KW immunosuppressive; sepsis.

XX Enterococcus faecalis V583.

XX

XX Location/Qualifiers

FT Region 6..30 /note= "Serum reactive epitope, claimed"

FT Region 36..42 /note= "Serum reactive epitope, claimed"

FT Region 143..157 /note= "Serum reactive epitope, claimed"

FT Region 176..197 /note= "Serum reactive epitope, claimed"

FT Region 202..209 /note= "Serum reactive epitope, claimed"

FT Region 216..233 /note= "Serum reactive epitope, claimed"

FT Region 226..269 /note= "Serum reactive epitope, claimed"

FT Region /note= "Serum reactive epitope, claimed"  
 FT 241..246 /note= "Serum reactive epitope, claimed"  
 FT 275..287 /note= "Serum reactive epitope, claimed"  
 FT 292..299 /note= "Serum reactive epitope, claimed"  
 FT 315..325 /note= "Serum reactive epitope, claimed"  
 FT 343..350 /note= "Serum reactive epitope, claimed"  
 FT 375..380 /note= "Serum reactive epitope, claimed"  
 FT 397..403 /note= "Serum reactive epitope, claimed"  
 FT 411..420 /note= "Serum reactive epitope, claimed"  
 FT 422..434 /note= "Serum reactive epitope, claimed"  
 FT 441..448 /note= "Serum reactive epitope, claimed"  
 FT 467..474 /note= "Serum reactive epitope, claimed"  
 FT 477..499 /note= "Serum reactive epitope, claimed"  
 FT 555..568 /note= "Serum reactive epitope, claimed"  
 FT 591..597 /note= "Serum reactive epitope, claimed"  
 FT 601..609 /note= "Serum reactive epitope, claimed"  
 FT 623..644 /note= "Serum reactive epitope, claimed"  
 FT 667..688 /note= "Serum reactive epitope, claimed"  
 FT 692..698 /note= "Serum reactive epitope, claimed"  
 FT 703..718 /note= "Serum reactive epitope, claimed"  
 FT 736..747 /note= "Serum reactive epitope, claimed"  
 FT 757..766 /note= "Serum reactive epitope, claimed"  
 FT 782..791 /note= "Serum reactive epitope, claimed"  
 FT 795..801 /note= "Serum reactive epitope, claimed"  
 FT 832..840 /note= "Serum reactive epitope, claimed"  
 FT 859..865 /note= "Serum reactive epitope, claimed"  
 FT Region /note= "Serum reactive epitope, claimed"  
 PN WO2004106367-A2.  
 PD 09-DEC-2004.  
 XX 26-MAY-2004; 2004WO-EP005664.  
 PR 30-MAY-2003; 2003EP-00450137.  
 PA (INTE-) INTERCELL AG.  
 PI Meinke A, Nagy E, Hanner M, Gelbmann D;  
 DR N-PSDB; ADVA16703.  
 XX WPI, 2005-039707/04.  
 PT Novel isolated nucleic acid molecule encoding hyperimmune serum reactive  
 PT antigen e.g., EP0020, EP0032, EP0062, EP0149, EP0253, EP0270,  
 PT EP0298 or its fragment, useful for producing vaccine against enterococcal  
 PT infection.  
 XX Claim 11, SEQ ID NO 450; 175pp; English.

XX The invention relates to a novel isolated nucleic acid molecule encoding  
 CC a hyperimmune serum reactive antigen e.g. SEQ ID 171-340, 357-372 or 425-  
 CC 476, or its fragment. The antigen of the invention demonstrates  
 CC antibacterial activities and may be useful for producing a pharmaceutical  
 CC preparation, particularly a vaccine, against enterococcal infection.  
 CC Conditions associated with bacterial infection which may be prevented or  
 CC treated include bacterial pharyngitis, scarlet fever, impetigo, rheumatic  
 CC fever, necrotizing fasciitis and sepsis in humans. The current sequence  
 CC is that of an Enterococcus faecalis V583 hyperimmune serum reactive  
 CC antigen protein of the invention.  
 XX  
 SQ Sequence 893 AA:  
 Query Match 6.7%; Score 119; DB 9; Length 893;  
 Best Local Similarity 19.4%; Pred. No. 0.38;  
 Matches 74; Conservative 58; Mismatches 119; Indels 130; Gaps 20;  
 QY 21 LETGLDN-----CNVANGLMIGVLIITPTDEGNFVID---DVTLDNITKIVD 69  
 DB 138 MKGGLSRVAKATVAVANVANDPVLPG-----KNFIDVSMNGDISVAEYOKIKS 186  
 QY 70 Y-IDGSDIDGDCMFTYGNPENEYNTIPNSQSYLSKSENSQITQIKRYVSCNYSRLRTK 128  
 DB 187 YGVTGVSVKLTGWTYV-NPYAAGQIRMAAGLAKYS---AVHSMYVSA--TAODEAR 240  
 QY 129 SEBAKVTTSQKVIISTONSINSRRVYNAIDATNFTDDELRTTKETRFENQSYTHKSS 188  
 DB 241 YF-AQAAANSQ-----LDKVTIMFNDAPPTLTNNGRNHANSVAFNQQLKALGYK 230  
 QY 189 TNSLYVHTWTI-----PSLIKQNRWEDYNNQWMAQSCY 225  
 DB 291 NDALVYKMWLTNGYIDTSAFGDRVWVAQYPTTPDS---SQMNNDHQAQWQSSQMTF 346  
 QY 226 KTGADGSESTRWLAAGSIFPPG--NYDGLMLNDIALSGMAH-----KSY----- 269  
 DB 347 -----PGLAVYEGRPDISMTYSFLLMGNSGPDLSKYVTNP 385  
 QY 270 -----NVDTGINQLSF-----TRILGFSWYTNISGLDGHAVITIDQ---YG 310  
 DB 386 GRVIMKNDPTFYODVAFRTPGWVRKKNVLTIGIR--YSSAGIPR---LVTDQGYLTA 439  
 QY 311 NKRYRILFHAGYENDPYLSSS 331  
 DB 440 NKDYVL--AAQSNIDLYFTTN 458  
 RESULT 13  
 ADH86812  
 ID ADH86812 standard; protein; 899 AA.  
 AC ADH86812;  
 XX  
 DT 22-APR-2004 (first entry)  
 XX  
 DE Enterococcus faecalis polypeptide #1292.  
 XX  
 KW Enterococcus faecalis infection; transcription regulatory element;  
 KM antibacterial.  
 OS Enterococcus faecalis.  
 XX  
 PN US6617156-B1.  
 XX  
 PD 09-SEP-2003.  
 XX  
 PF 13-AUG-1998; 98US-00134000.  
 XX  
 PR 15-AUG-1997; 97US-0055778P.  
 XX  
 PA (DOUC/) DOUCETTE-STAMM L A.  
 XX (BUSH/) BUSH D.  
 XX





Db 115 YTTIQVTVMQNTINEYTFEASVNDTIAFSTSGGGQGLPEPKTYKIGPVWEDVDKDG 174  
 QY 145 TQNSINSRVINAIDATNTDDELRTTKETREFNOSYTHSKSTNSL-YVHTTIRSL 203  
 Db 175 IQNTNDEKLSNVLTLYTPDG---YTSKVRTEDEDERKYPDGKLNLTLYKIPETP--- 228  
 QY 204 KLGQWRMEDY-----NNGWTMAQSCYVTKGADGSESTRWLAAGSIFP 246  
 Db 229 -----EGYTPPLKSGCTNPALDSEGSVWVTINGDDMTIDSGFYQTKYSLGAVWY 281  
 QY 247 PGVYDGLMDNDIALSGMAHKSYNVDTGINQLSFTRIIGKGFVWYNIISGLDRGHAVIIT 306  
 Db 282 DTNMDGIGDDEKISGV--KVTLKDENGNIISTTTDENG---KYQPDNLNSGNYIVHF 336  
 QY 307 DQ 308  
 Db 337 DK 338

## RESULT 15

AEB91631  
 ID AEB91631 standard; protein; 773 AA.

AC AEB91631;  
 XX  
 DT 20-OCT-2005 (first entry)  
 XX

DE Microbial pathogen adhesin protein sequence, SEQ ID NO:341.

XX  
 KM algorithm; adhesin; pharmaceutical; vaccine; drug screening;  
 KM bordetella pertussis infection; antibacterial; pneumonia;  
 KM anti-inflammatory; respiratory-gen; gastric ulcer; anticler;  
 KM gastrointestinal-gen; urinary tract infection; antimicrobial; uropathic.

OS Shigella flexneri.

PN MO2005076010-A2.

PD 18-AUG-2005.

PF 07-FEB-2005; 2005WO-IN000037.

PR 06-FEB-2004; 2004IN-DE000173.

PR 20-JUL-2004; 2004US-0589227P.

XX (COUL ) COUNCIL SCT & IND RBS SOUTH AFRICA.

PA Sachdeva G, Kumar K, Jain P, Brahmachari SK, Ramchandran S;

PI WPI; 2005-597835/61.

DR  
 XX  
 PT Computational method for identifying adhesin and adhesin like molecules,  
 PT comprises computing sequence-based attributes of protein sequences using  
 PT neural network software and training an artificial neural network.

PS Claim 17; SEQ ID NO 341; 402pp; English.

CC The present invention relates to a computational method (M1) for  
 CC identifying adhesin and adhesin-like proteins, by computing the sequence-  
 CC based attributes of protein sequences using five attribute modules of a  
 CC neural network software, training an artificial neural network (ANN) for  
 CC each of the computed five attributes, and identifying the adhesin and  
 CC adhesin-like proteins having probability of being an adhesin (Pad) as  
 CC equal or greater than 0.51. Also claimed is a set of 274 annotated genes  
 CC encoding adhesin and adhesin-like proteins, having 274 fully defined 162-  
 CC 1151 base pair (SEQ ID NO: 385-658) sequences; a set of 105 hypothetical  
 CC genes encoding adhesin and adhesin-like proteins, having 105 fully  
 CC defined 306-15876 base pairs (SEQ ID NO: 659-763); a set of 279 annotated  
 CC adhesin and adhesin-like proteins, having 279 fully defined 53-3716 base  
 CC pair (SEQ ID NO: 1-279); a set of 105 hypothetical adhesin and adhesin-  
 CC like proteins, having 105 fully defined 106-5291 base pair (SEQ ID NO:  
 CC 280-384) sequences; and a fully connected multilayer feed forward ANN (1)  
 CC based on (M1). (M1) is useful for identifying adhesin and adhesin-like

CC proteins, of therapeutic potential, and identifying and short-listing  
 CC proteins for further testing in development of new vaccine formulations  
 CC to eliminate diseases caused by various pathogenic organisms. (M1) is  
 CC useful for identifying putative adhesins that are important in drug  
 CC discovery and preventing therapeutics for whooping cough, pneumonia,  
 CC gastric ulcer and urinary tract infections. (M1) identifies adhesins from  
 CC distantly related organisms, and from bacteria belonging to a wide  
 CC phylogenetic spectrum. (M1) is capable of predicting adhesive nature of  
 CC unique proteins. The present sequence is a microbial pathogen adhesin  
 CC protein sequence.

SQ Sequence 773 AA;

Query Match 6.6%; Score 118; DB 9; Length 773;

Best Local Similarity 22.7%; Pred. No. 0.37;

Matches 83; Conservative 55; Mismatches 151; Indels 76; Gaps 18;

QY 2 TLREDSMDWTGVSTFNVILET-----GLDNCNIYANGNMGVITINPTDDEG 51  
 Db 76 SLDEESTAD-TGSNNENNAIAKMDAGEITTHGTESSAAYAN-----GTVYKAGDTLDYT 129  
 QY 52 N---FVDIDVTLDNLIKIDYIDGSDIDGSDGWFYTGPNF-----YN---TIPNSQ 98  
 Db 130 NASVTLTDVDTTHGDNNAHAIAARQGT-VSFNGEITTTGPDAAIAKIYNGTIVTLKNTS 188  
 QY 99 SYS-----LKSEBSQITQIRYVSCSTSLRT-----KPSAKYTTTSGVIST 144  
 Db 189 AVAHQSGGLVLESSINGQ---SATVDILSGSLRSANIELYHKDETSNVTITDSEVSA 244  
 QY 145 TQNSINSRVINAIDATNTDDELRTTKETREFNOSYTHSKSTNSLYVHTTIRSLK 204  
 Db 245 ADVFINNIKHL-TVADTN---SKITGSANISTDDWHTYLSLSDS---TWDIDRDSF 236  
 QY 205 LQWRMEDYNNGTMAQSCYVTKGADGS-ESTRWLAAGSIFPPGVYDGLMDNDIALSG 263  
 Db 297 VSN-----LTVNSFTVYISRADRDVEPRLTITENYV--GNNGLVHLRTELDN 345  
 QY 264 MAHKSYNVDGINQLSFTRIIGKGFVWYNIISGLDRGHAVIITDQGN---KYRLIFHA 319  
 Db 346 SATDKVINGNTSGTTRKVTNAGSGGAYTLNIE---IISVEGSGNGBFIDSRIFAG 401  
 QY 320 GYENS 324  
 Db 402 AYEYS 406

Search completed: January 30, 2006, 09:46:03  
 Job time : 28.6283 secs

**THIS PAGE LEFT BLANK**

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2006, 09:42:59 ; Search time 7.48879 Seconds  
(without alignments)  
3687.335 Million cell updates/sec

Title: US-09-889-874a-22

Perfect score: 1784  
Sequence: 1 FTLRSDMSMDWTGVSTFNV.....ILFHAGYNSDPYLSISIV 334

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued Patents AA: \*  
1: /cgn2\_6/ptodaca/1/1aa/5 COMB pep: \*  
2: /cgn2\_6/ptodaca/1/1aa/6 COMB pep: \*  
3: /cgn2\_6/ptodaca/1/1aa/H COMB pep: \*  
4: /cgn2\_6/ptodaca/1/1aa/PTCUS COMB pep: \*  
5: /cgn2\_6/ptodaca/1/1aa/RB COMB pep: \*  
6: /cgn2\_6/ptodaca/1/1aa/backfile1.pcp: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	126	7.1	930	US-09-134-001C-5314	Sequence 5314, Ap
2	126	7.1	930	US-09-386-962C-10	Sequence 10, Appl
3	124	7.0	1092	US-09-147-405B-15	Sequence 15, Appl
4	119	6.7	899	US-09-134-000C-4697	Sequence 4697, Ap
5	115.5	6.5	1411	US-10-080-505-17	Sequence 17, Appl
6	114	6.4	360	US-09-107-433-5178	Sequence 5178, Ap
7	114	6.4	376	US-09-583-110-5138	Sequence 13, Appl
8	112.5	6.3	1436	US-10-080-505-13	Sequence 13, Appl
9	107.5	6.0	324	US-09-830-230A-418	Sequence 418, Ap
10	107.5	6.0	343	US-09-830-230A-417	Sequence 417, Ap
11	106	5.9	258	US-09-251-645-5	Sequence 5, Appl
12	105	5.9	1430	US-09-008-172-2	Sequence 2, Appl
13	105	5.9	1430	US-09-210-361-6	Sequence 6, Appl
14	105	5.9	1430	US-09-740-274-6	Sequence 6, Appl
15	104	5.8	1306	US-09-538-092-330	Sequence 330, Ap
16	104	5.8	1589	US-09-543-681A-4998	Sequence 4998, Ap
17	103.5	5.8	582	US-09-147-405B-13	Sequence 13, Appl
18	103.5	5.8	583	US-09-147-405B-11	Sequence 11, Appl
19	103	5.8	494	US-09-543-681A-4773	Sequence 4773, Ap
20	103	5.8	607	US-08-409-995-6	Sequence 6, Appl
21	103	5.8	607	US-08-685-467-6	Sequence 6, Appl
22	103	5.8	607	US-08-913-942-6	Sequence 6, Appl
23	103	5.8	607	US-09-684-707-6	Sequence 6, Appl
24	103	5.8	1395	US-10-080-505-7	Sequence 7, Appl
25	103	5.8	1912	US-08-409-995-4	Sequence 4, Appl
26	103	5.8	1912	US-08-685-467-4	Sequence 4, Appl
27	103	5.8	2353	US-09-377-155-33	Sequence 33, Appl

28	103	5.8	2353	2	US-08-913-942-4	Sequence 4, Appl
29	103	5.8	2353	2	US-09-669-974-33	Sequence 33, Appl
30	103	5.8	2353	2	US-09-797-862-33	Sequence 33, Appl
31	103	5.8	2353	2	US-09-684-707-4	Sequence 4, Appl
32	103	5.8	2353	2	US-09-268-347-47	Sequence 47, Appl
33	103	5.8	2411	2	US-09-268-347-36	Sequence 36, Appl
34	102.5	5.7	1963	2	US-09-583-110-5243	Sequence 5243, Ap
35	102.5	5.7	1967	2	US-09-107-433-4883	Sequence 4883, Ap
36	101.5	5.7	545	2	US-09-604-957-4	Sequence 4, Appl
37	101.5	5.7	545	2	US-09-995-749A-10	Sequence 10, Appl
38	101.5	5.7	1224	2	US-09-901-572A-4	Sequence 4, Appl
39	101	5.7	1434	2	US-10-080-505-9	Sequence 9, Appl
40	101	5.7	1457	2	US-09-673-896-4	Sequence 4, Appl
41	101	5.7	1457	2	US-09-303-518D-650	Sequence 650, Ap
42	100.5	5.6	829	2	US-09-514-599-6	Sequence 6, Appl
43	100.5	5.6	829	2	US-09-996-024-6	Sequence 6, Appl
44	100.5	5.6	862	2	US-09-346-237-1	Sequence 1, Appl
45	100.5	5.6	1781	2	US-09-395-749A-2	Sequence 2, Appl

#### ALIGNMENTS

```
RESULT 1
US-09-134-001C-5314
; Sequence 5314, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucelte-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134, 001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064, 964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055, 779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5314
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5314

Query Match 7.1%; Score 126; DB 2; Length 930;
Best Local Similarity 21.0%; Pred. No. 0.0057;
Matches 78; Conservative 51; Mismatches 141; Indels 102; Gaps 16;

QY 9 SDWTGVSTFVILETGLDNCNIYANGLMNG--VIINFTDDEGNFVDIDVT----- 60
DB 436 SMFTNIDTKNTHVEQ-----TIYINPLRYSAKETNNISGNGDGSIT-IDSTTIKRVK 489
QY 61 -----LNDIKIKVDYIDSGIDSGWFFYTGPNENVT-----IPNS 97
DB 490 VEGNQMLPDSNRKYDSEYEDVT-NDYQAQGNNDVNIQNDISPIYIKVSKIDPNK 548
QY 98 QSYSLKSENSQITQIKRYVSCNTRSLRTPSARV--TTTSG-----KVISI-- 144
DB 549 DDYTTIQVTFMQTTINFY-----TGEPRYASDNTIAFSTSGGCGDLPPEKTYKIGD 603
QY 145 -----TQNSINSSRVVAINAINTPTDELATTTETRENGSYTHKSTNSL-Y 193
DB 604 YWEDVDKQGIQNTNKEPLSNVLTLYTPDG---TSKSVRTDEGKQOPDGLKNGLTY 660
QY 194 VHTTTPRSIKLQNMWEDY-----NNGMTWAQSCYYTIGADGSEST 236
DB 661 KITFETP-----EGTPTLKSGTNPALDSRGNSWVITINQDDMTIDSGFYQTP 710
QY 237 RWLAQSIRPPGNYDGLMTNDIALSGMAHKSYNVDTGINQLSPYRIIGKGSWVYNISG 236
DB 711 KXSLGNYVWYDNTKQIGDDEKIGISGV--KVLTKDENGNIISTTTTDENG--KYQFDN 765
```

QY 297 LDRGHAVIIDIQ 308  
| : : : :  
Db 766 LNSGNVIHFDPK 777

## RESULT 2

US-09-386-962C-10  
; Sequence 10, Application US/09386962C  
; Patent No. 6635473  
; GENERAL INFORMATION:  
; APPLICANT: FOSTER, Timothy  
; TITLE OR INVENTION: POLYPEPTIDES AND POLYNUCLEOTIDES FROM COAGULASE-NEGATIVE STAPHYLOCOCCUS  
; FILE REFERENCE: P06335US2/BAS  
; CURRENT APPLICATION NUMBER: US/09/386,962C  
; PRIOR FILING DATE: 1999-08-31  
; PRIOR APPLICATION NUMBER: 60/098,443  
; PRIOR FILING DATE: 1998-08-31  
; PRIOR APPLICATION NUMBER: 60/117,119  
; PRIOR FILING DATE: 1999-01-25  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 10  
; LENGTH: 930  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-386-962C-10

Query Match 7.1%; Score 126; DB 2; Length 930;

Best Local Similarity 21.0%; Pred. No. 0.0057;

Matches 78; Conservative 51; Mismatches 141; Indels 102; Gaps 16;

QY 9 SDMTGVSTFNVILETGLDNCNIYANGLMIG--VIINITPTDDEGNFVDIDVT-----60  
| : : : : :  
Db 436 SMFTNIDTKHHTVEQ-----TIYINPLRYSAKETNIVISGNGDEGSTI-IDDSIIIRKYK 489  
| : : : : :  
QY 61 -----LNDNIXIVDYIDSGSDGWFYTGPNENYNT-----IPNS 97  
| : : : : :  
Db 490 VGDQNLPPSNRIYDYSEYEDVT-NDDYAQLGNNDVNIINFGNIDSEPIIKIVISKYDPNK 548  
| : : : : :  
QY 98 QSYSLKSENSQITQIKRYVSCNTSRLRTKSPSAKV--TTTSG-----KVISI-- 144  
| : : : : :  
Db 549 DDTTIOQTYMTQTTINEY-----TGEFRITASYNNTIAFSTSSGQGGDLPEPKTYKIGD 603  
| : : : : :  
QY 145 -----TONSINSSRVVINAIDATNFTDELRTTKETREPNOSYTSKSTNSL-Y 193  
| : : : : :  
Db 604 YWVEDVDKQIGQNTNDEKPLSNVLVLTYPDG---TSKSVRTDEDEKYQFDGKNGLTLY 660  
| : : : : :  
QY 194 VHTWTIPRSKLQWNRWEDY-----NNGWTAQSCYYKTGADGSEST 236  
| : : : : :  
Db 661 KITFETP-----EGYTPILKHSCTNPALDSEGNVWVTIINGDDMTIDSGFYQTP 710  
| : : : : :  
QY 237 RWLAAGSIFPPGNYDGLMLNDIALSGMAHKSYNVDGINQLSFTRIIGKGFVWVNI 296  
| : : : : :  
Db 711 KYSIGNTVWYDTNKDQIGSDDEKGISGV--KYLKDKNGNIISTTTDENG---KYQPDN 765  
| : : : : :  
QY 297 LDRGHAVIIDIQ 308  
| : : : : :  
Db 766 LNSGNVIHFDPK 777

## RESULT 3

US-09-147-405B-15  
; Sequence 15, Application US/09147405B  
; Patent No. 6733758  
; GENERAL INFORMATION:  
; APPLICANT: Guss, Bengt  
; APPLICANT: Nilsson, Martin  
; APPLICANT: Frykberg, Lars  
; APPLICANT: Flock, Jan-Ingemar  
; APPLICANT: Lindberg, Martin  
; TITLE OR INVENTION: Fibrinogen Binding Protein Originating from  
; TITLE OF INVENTION: Coagulate-Negative Staphylococcus  
; FILE REFERENCE: guss 09/147405

; CURRENT APPLICATION NUMBER: US/09/147,405B  
; CURRENT FILING DATE: 1999-04-11  
; PRIOR APPLICATION NUMBER: PCT/SE97/10191  
; PRIOR FILING DATE: 1997-06-18  
; PRIOR APPLICATION NUMBER: SE 9602496-3  
; PRIOR FILING DATE: 1996-06-20  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 15  
; LENGTH: 1092  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-147-405B-15

Query Match 7.0%; Score 124; DB 2; Length 1092;

Best Local Similarity 21.0%; Pred. No. 0.011;

Matches 78; Conservative 50; Mismatches 142; Indels 102; Gaps 16;

QY 9 SDMTGVSTFNVILETGLDNCNIYANGLMIG--VIINITPTDDEGNFVDIDVT-----60  
| : : : : :  
Db 439 SMFTNIDTKHHTVEQ-----TIYINPLRYSAKETNIVISGNGDEGSTI-IDDSIIIRKYK 492  
| : : : : :  
QY 61 -----LNDNIXIVDYIDSGSDGWFYTGPNENYNT-----IPNS 97  
| : : : : :  
Db 493 VGDQNLPPSNRIYDYSEYEDVT-NDDYAQLGNNDVNIINFGNIDSEPIIKIVISKYDPNK 551  
| : : : : :  
QY 98 QSYSLKSENSQITQIKRYVSCNTSRLRTKSPSAKV--TTTSG-----KVISI-- 144  
| : : : : :  
Db 552 DDTTIOQTYMTQTTINEY-----TGEFRITASYNNTIAFSTSSGQGGDLPEPKTYKIGD 606  
| : : : : :  
QY 145 -----TONSINSSRVVINAIDATNFTDELRTTKETREPNOSYTSKSTNSL-Y 193  
| : : : : :  
Db 607 YWVEDVDKQIGQNTNDEKPLSNVLVLTYPDG---TSKSVRTDEDEKYQFDGKNGLTLY 663  
| : : : : :  
QY 194 VHTWTIPRSKLQWNRWEDY-----NNGWTAQSCYYKTGADGSEST 236  
| : : : : :  
Db 664 KITFETP-----EGYTPILKHSCTNPALDSEGNVWVTIINGDDMTIDSGFYQTP 713  
| : : : : :  
QY 237 RWLAAGSIFPPGNYDGLMLNDIALSGMAHKSYNVDGINQLSFTRIIGKGFVWVNI 296  
| : : : : :  
Db 714 KYSIGNTVWYDTNKDQIGSDDEKGISGV--KYLKDKNGNIISTTTDENG---KYQPDN 768  
| : : : : :  
QY 297 LDRGHAVIIDIQ 308  
| : : : : :  
Db 769 LNSGNVIHFDPK 780

## RESULT 4

US-09-134-000C-4697  
; Sequence 4697, Application US/09134000C  
; Patent No. 6617156  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OR INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 032796-032  
; CURRENT APPLICATION NUMBER: US/09/134,000C  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/055,778  
; PRIOR FILING DATE: 1997-08-15  
; NUMBER OF SEQ ID NOS: 6812  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4697  
; LENGTH: 899  
; TYPE: PRT  
; ORGANISM: Enterococcus faecalis  
US-09-134-000C-4697

Query Match 6.7%; Score 119; DB 2; Length 899;

Best Local Similarity 19.4%; Pred. No. 0.025;

Matches 74; Conservative 58; Mismatches 119; Indels 130; Gaps 20;

QY 21 LETGLDN-----CNIYANGLMIGVIINITPTDDEGNFVDID----DVTLNDNIXIVD 69

Db 144 MKKGLSRVAVATVAVVANDPMLPG-----KNTIDVSNMGDISVAEYKIKS 192  
Qy 70 Y-IDSGDIDGSDGFTGNENETITPNSOSYSLKSENSQTOIKRYVSCSTSLRTRK 128  
Db 193 YGTGVSVKLTETGTVV-NPYAAGQIRNAKAGIKYS---AVHYSYSA-A-TAODEAR 246  
Qy 129 SPKAYTTTSGKYSTIGNSINSRVYINADATNFDDDELRTTKETRFENOSYTSKKS 188  
Db 247 YF-AQAAANSQ-----LDKNTWFDADPTLTNNRMAHANSVAFNOOLKALGYK 296  
Qy 189 TNSLYVHTWTI-----PRSLKQNRWEDYNNGTMAOSCY 225  
Db 297 NDALVYKMWLTNGYIDTSAFGDRVVAQYPTPOS---SMQNNHDGAMQMSQMYF 352  
Qy 226 KTGADGSESTRMLAGSTPPG--NYDGLMNDIALSGMAH-----KSY----- 269  
Db 353 -----PGLANYEGRPFDISMTYSNPLMGNSSGPDLSKYTYTNP 391  
Qy 270 -----NVDGINGLSP-----TRIGKFSVNVNIGLDRGHAVIIDQ--YG 310  
Db 392 GRVIMKNDTFYQDVARTGMRVKKNTLVTKIGIR--YSSAGIPR---LVTDQGYLTA 445  
Qy 311 NKXRIIFHAGYENSDEPLSSS 331  
Db 446 NKDYVL--AAQSNIDLYFTTN 464

RESULT 5  
US-10-080-505-17  
Sequence 17, Application US/10080505  
Patent No. 6676948  
GENERAL INFORMATION:  
APPLICANT: St. Gene, Joseph W.  
TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS  
FILE REFERENCE: A-59941-1/RT/DCF/DHR  
CURRENT APPLICATION NUMBER: US/10/080,505  
CURRENT FILING DATE: 2002-02-22  
PRIOR APPLICATION NUMBER: US 08/296,791  
PRIOR FILING DATE: 1994-10-25  
PRIOR APPLICATION NUMBER: US 09/839,996  
PRIOR FILING DATE: 2001-04-20  
NUMBER OF SEQ ID NOS: 58  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 17  
LENGTH: 1411  
TYPE: PRT  
ORGANISM: Haemophilus influenzae  
US-10-080-505-17

Query Match 6.5%; Score 115.5; DB 2; Length 1411;  
Best Local Similarity 20.8%; Pred. No. 0.11;  
Matches 85; Conservative 48; Mismatches 131; Indels 145; Gaps 21;  
Qy 4 REDMSDMTGV-----TENVILEGLNCNIYANGLMIGVITIT----- 45  
Db 554 RDLAFNGMFGDKDXTKATGLANTYNP--NKNHFLSGGTYLKG--NITDGGTIV 607  
Qy 46 -----PTDDEGNFVD-----IDVTLLNDNIKYDY-IDG-----SDIDG 78  
Db 608 FSGRPTTHAVNHLRLNELGRKGEVVIDDMVIRTKAENFQJKGSGTVVSRVNSIEG 667  
Qy 79 SDGFTYGNPN-EVNTIPNSOSYSLKSENSQTOIKRYVSCSTSLRTRKTSFAKTTT 137  
Db 668 N-WTISNNANATGVVPOONTICTRSDWGLT-----TCKTVNLTDKVIDISPTTQ 719  
Qy 138 SGKYSTIONS-----INSGRVYIN-----AIDATN- 163  
Db 720 INGSINNNATVNIHGLAKINGVTLNHSQFTLSNNAOTGNIQLSNNAATVDANL 779  
Qy 164 -----FTDDELRTTKETRFENOSYTSKKSSTNSLYVHTTIPRSLKQNRWEDYNNGT 218  
Db 780 NGVHLLDSAQFSLKNSHFSHQ-IQGDQDTVTLENATWTPSDTTLQNL---TLNNSV 835

Qy 219 MAOSCY-----YKTGADGSESTRW---LAAGSIFPGANYDL-----WLDN 257  
Db 836 TANSAYSASNNAPRRRLBETFTPTSEHFNLTUVNGKLSGCTPFTSSLRKYSD 895  
Qy 258 DIALSGMAKSYNV---DTG-----INQSLFTRIIGKFSWYVNSIGLD 298  
Db 896 KIKLSNDAGDYTLAVRDTGKEPVTLEQLTL-----IEGLD 931

RESULT 6  
US-09-107-433-5178  
Sequence 5178, Application US/09107433  
Patent No. 6800744  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGNOSIS  
THERAPEUTICS  
NUMBER OF SEQUENCES: 5206  
CORRESPONDENCE ADDRESS:  
ADDRESSER: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: <Unknown>  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: <Unknown>  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,433  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085131  
FILING DATE: May 12, 1998  
APPLICATION NUMBER: 60/051553  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Arinello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-011  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 5178:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 360 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULAR TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Streptococcus pneumoniae  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: (B) LOCATION 1...360  
SEQUENCE DESCRIPTION: SEQ ID NO: 5178:  
US-09-107-433-5178

Query Match 6.4%; Score 114; DB 2; Length 360;  
Best Local Similarity 25.6%; Pred. No. 0.019;  
Matches 46; Conservative 33; Mismatches 73; Indels 28; Gaps 8;  
Qy 31 YANGLMNIGVILN--TTPDDEGNFVDIDVTLLNDNIKYDYIDG-----SDIDG 78  
Db 104 YKRAHQAISVTLARKGALPIINENDSVVIDEVKGDMDTLSAQVAAWQADLVLTDVDS 163  
Qy 79 SDGFTYGNENETITPNSOSYSLKSENSQTOIKRYVSCSN-TSLRTRKTSFAKTTT 137  
Db 164 -----LVYGNP---NSDPRAKRLRIETINREIIDMAGAGASSNGTGMLTKIKAAATATE 216

Qy 138 SGKVISTONSINSRVINAIDATNFTDDE---LRTTKETRENOSYTHSKSSTNSLYV 194  
Db 217 SGVPEVYIC-SSLSKD---SWIEBAETEDSGSYFAQEKIKTKOKWLAFAVQSQSSIMV 271

## RESULT 7

US-09-583-110-5138  
; Sequence 5138, Application US/09583110  
; Patent No. 6699703  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al.  
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus  
; FILE REFERENCE: PAT400-07A  
; CURRENT APPLICATION NUMBER: US/09/583,110  
; CURRENT FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/107,433  
; PRIOR FILING DATE: 1998-06-30  
; PRIOR APPLICATION NUMBER: US 60/085,131  
; PRIOR FILING DATE: 1998-05-12  
; PRIOR APPLICATION NUMBER: US 60/051,553  
; PRIOR FILING DATE: 1997-07-02  
; NUMBER OF SEQ ID NOS: 5322  
; SEQ ID NO 5138  
; LENGTH: 376  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-583-110-5138

Query Match 6.4%; Score 114; DB 2; Length 376;  
Best Local Similarity 25.6%; Pred. No. 0.02;  
Matches 46; Conservative 33; Mismatches 73; Indels 28; Gaps 8;

Qy 31 YANGLMIGVIIN--ITPTDEGNFVDIDVTLNDNIKIYDIIG-----SDIDG 78  
Db 120 YKNAHQALSVILNKGAFIINENDSVVIDEYKVGNDITLSAQVAMVQADLLVLTVDG 179  
Qy 79 SDGAFYGNPEVNTIPNSQSYSLKSENQITQIKRVSGSN--TSRLRTSFSKAYTTT 137  
Db 180 ---LYTGNP---NSDPKAKLERIETTNREIIDPAGAGSSNGGKWTIKKATYATE 232  
Qy 138 SGKVISTONSINSRVINAIDATNFTDDE---LRTTKETRENOSYTHSKSSTNSLYV 194  
Db 233 SGVPEVYIC-SSLSKD---SWIEBAETEDSGSYFAQEKIKTKOKWLAFAVQSQSSIMV 287

## RESULT 8

US-10-080-505-13  
; Sequence 13, Application US/10080505  
; Patent No. 6676948  
; GENERAL INFORMATION:  
; APPLICANT: St. Gene, Joseph W.  
; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS  
; FILE REFERENCE: A-59941-1/RFA/DCF/DHR  
; CURRENT APPLICATION NUMBER: US/10/080,505  
; CURRENT FILING DATE: 2002-02-22  
; PRIOR APPLICATION NUMBER: US 08/296,791  
; PRIOR FILING DATE: 1994-10-25  
; PRIOR APPLICATION NUMBER: US 09/839,996  
; PRIOR FILING DATE: 2001-04-20  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 13  
; LENGTH: 1436  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-10-080-505-13

Query Match 6.3%; Score 112.5; DB 2; Length 1436;  
Best Local Similarity 21.1%; Pred. No. 0.22;  
Matches 64; Conservative 39; Mismatches 103; Indels 97; Gaps 14;

Qy 4 REDSMSDMTGVSTFN-----VILETGDNQNIYANGLMIGVIINIT----- 45  
Db 570 KEIAYNMCFEETDENKNGNRLNIYKPTTEDRTLILSGGNNLNG--NITQEGSTLVFSG 626  
Qy 46 -PTDDEGNFVD-----IDVTLNDNIKIYDI- IDG-----SDIDGSDG 81  
Db 627 RPTPHAYNHLNRDNEIGRPOGEVVIDDMITRTFKENFOIKGSAAVSRVSSIEGN-- 684  
Qy 82 WFTYGNPN-EYNTIPNSQSYSLKSENQITQIKRVSGSNTSRLRTKPSAKYTTTSGK 140  
Db 685 WTVSNNAAPAGVAPNQNTICTRSDDTGLT-----TCKRTVDLTDTKVINSIPTQING 738  
Qy 141 VISITONS-----INSRVIVIN-----AIDATN----- 163  
Db 739 SIMLTDATYNIHGLAKNGVTLNHSQFTLSNNATQGNIQLSNANATVDNANLGN 798  
Qy 164 -FTDDELRTTKETRENOSYTHSKSSTNSLYVHTTTPSLKQKRMEDYNGGWTWAO 221  
Db 799 VHLTDSAQFSLKNSHFHQ-IOGDKDTVTLENATWTMPSDATLQNT---TLNNSVTTLN 854  
Qy 222 SCY 224  
Db 855 SAY 857

## RESULT 9

US-09-830-230A-418  
; Sequence 418, Application US/09830230A  
; Patent No. 6902893  
; GENERAL INFORMATION:  
; APPLICANT: Human Genome Sciences, Inc.  
; TITLE OF INVENTION: Lyme Disease Vaccine  
; FILE REFERENCE: PB481US  
; CURRENT APPLICATION NUMBER: US/09/830,230A  
; CURRENT FILING DATE: 2001-09-27  
; PRIOR APPLICATION NUMBER: PCT/US98/12718  
; PRIOR FILING DATE: 1998-06-18  
; PRIOR APPLICATION NUMBER: PCT/US98/12718  
; PRIOR FILING DATE: 1997-09-03  
; PRIOR APPLICATION NUMBER: 60/057,483  
; PRIOR FILING DATE: 1997-09-03  
; PRIOR APPLICATION NUMBER: 60/053,344  
; PRIOR FILING DATE: 1997-07-22  
; PRIOR APPLICATION NUMBER: 60/053,377  
; PRIOR FILING DATE: 1997-07-22  
; PRIOR APPLICATION NUMBER: 60/050,359  
; PRIOR FILING DATE: 1997-06-20  
; NUMBER OF SEQ ID NOS: 756  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 418  
; LENGTH: 324  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-830-230A-418

Query Match 6.0%; Score 107.5; DB 2; Length 324;  
Best Local Similarity 18.4%; Pred. No. 0.067;  
Matches 44; Conservative 38; Mismatches 80; Indels 77; Gaps 9;

Qy 14 VSTFNVLFTGLDNCNIYANGLMIGVIINTPTDDEGNFVDIDVTLNDNIKIYDIYIDG 73  
Db 122 INYKFTLNGIDILN-----NEKGNFYNALSLIED---VQDY--- 158  
Qy 74 SDIDSGWPFYTGPNNEYTIP-----NSQSYSLKSENQITQ 112  
Db 159 -----DSYIFY---KFLSIPRAHLKIDSNDYFNVTKINYNPPEVVRNIGDLIQD 210  
Qy 113 IKRVYSCSNTSRL-----RTKSPSAKYTTTSGKVISTONSINSRVINAIDATNFTDD 167  
Db 211 VKNFVLGNTSKLINIRDKNPFQSWDOKGK-----SNSINTNSFLTWTIRLGGRKN 265  
Qy 168 ELRTTKETRENOSYTHSKSSTNSLYVHTW-----TIPSLKQKRMEDYNGGWTW 219  
Db 266 GIQPAKHLSDSDISYLESRGWDHTHEWTFVKRIYVPOD-----PEINNGWTW 316

```
RESULT 10
US-09-830-230A-417
; Sequence 417, Application US/09830230A
; Patent No. 6902893
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Lyme Disease Vaccines
; FILE REFERENCE: P8481US
; CURRENT APPLICATION NUMBER: US/09/830,230A
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: PCT/US98/12718
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/057,483
; PRIOR FILING DATE: 1997-09-03
; PRIOR APPLICATION NUMBER: 60/053,344
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/053,377
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/050,359
; PRIOR FILING DATE: 1997-06-20
; NUMBER OF SEQ ID NOS: 756
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 417
; LENGTH: 343
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-830-230A-417

Query Match          6.0%; Score 107.5; DB 2; Length 343;
Best Local Similarity 18.4%; Pred. No. 0.073;
Matches 44; Conservative 38; Mismatches 80; Indels 77; Gaps 9;

QY 14 VSPFNVLEGLDNCNIVANGLMIGYIINTPPDDEGNFVDIDVTLNDNIKVDYIDG 73
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 141 INYKFLINNGIDILN-----NEKGNFYMLALSLIED---VQDY--- 177
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 74 SDIDSGDWFTYGNPNENYNTIP-----NSQSYLSKSNISQITQ 112
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 178 -----DESYRY--KXFLSLPRALMLKIDSDYNNVTKINYNFNNPFFVYRNGLDILQD 229
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 113 IKRYVSCNSTR-----RTKSPSAKYTTTSGKVISITONSINSSRVVINAIDATNFTD 167
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 230 VKQNVLSGNTSKLINDKNNFFIQSWDQKGGK-----SNSINNSFLTITMIRLGGRRKN 284
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 168 ELRTKTRFBNOSYTHSKSTNSLYVHTW-----TTPRSKLQNMWRMEDYNNNGWTW 219
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 285 GIQFAKHLBADSSDDISYLSRGMWDHHEWYFYFKRIYVPRD-----PEINNGWTW 335
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
US-09-251-645-5
; Sequence 5, Application US/09251645
; Patent No. 6281413
; GENERAL INFORMATION:
; APPLICANT: Kramer, Vance C.
; APPLICANT: Morgan, Michael K.
; APPLICANT: Anderson, Arne R.
; APPLICANT: Hart, Hope
; APPLICANT: Warren, Gregory W.
; APPLICANT: Dunn, Martha
; APPLICANT: Chen, Jeng S.
; TITLE OF INVENTION: NOVEL INSECTICIDAL TOXINS FROM PHOTORHABDUS LUMINESCENS
; FILE REFERENCE: CGC1963/A
; CURRENT APPLICATION NUMBER: US/09/251,645
; CURRENT FILING DATE: 1999-02-17
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 5
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
```

```
US-09-251-645-5

Query Match          5.9%; Score 106; DB 2; Length 258;
Best Local Similarity 21.3%; Pred. No. 0.066;
Matches 57; Conservative 44; Mismatches 96; Indels 70; Gaps 13;

QY 29 NIYANGLMIGYIINTPTD-DEGNFVD-----IDVTLNDNIKIVYID 72
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 16 NPYANGRHQCMVATSVLKQYTRGDWTKALSLAEKRSIOVALSBSLIDQLKM----- 70
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 73 GSDIDSGDWFTYGNPNENY-----NTIPNSQSY-----SLKSNISQITQ----- 112
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 71 -----PSGWTITDARKFDLGLNGVHADAFIDQVTDRAQDCCCTNENYQNSVKSVP 124
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 113 -IKRYVSCNSTRLKRFKSPAKYT--TTSKVISITONGIN-----SRVYVINAIDATNFT 165
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 125 IIRYVSSNRTS---TEYLMAKMTFBDTGKRLTITNMSVGDEVPFSKVLKIAAPYAIN 181
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 166 DDELRTTKETRFBNOSYTHSKSTNS--LYVHTWTIPRSKLQNMWRMEDYNNNGWTWAQSC 223
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 182 TNOLEHINITLFDKTEBPT-KSDTHHQIITLYYKMTLPYHLRILEGNDSTVNR----- 232
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 224 YKRTGADGSGESTRWLAAGSIPPPGNV 250
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 233 IYVLGKEPSND--RFLTRGRVFRKGTG 257
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 12
US-09-008-172-2
; Sequence 2, Application US/09008172
; Patent No. 6127602
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; FILE REFERENCE: 0358D
; CURRENT APPLICATION NUMBER: US/09/008,172
; CURRENT FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-008-172-2

Query Match          5.9%; Score 105; DB 2; Length 1430;
Best Local Similarity 22.5%; Pred. No. 1.1; Indels 110; Gaps 19;
Matches 82; Conservative 40; Mismatches 133;

QY 5 EDMSDPMWGVSTFNVILETGLDNCN--IYANGLMIGYIIN--ITPTDEGNF-----VD 55
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 411 EDNNS-----GVDPLANDIDNSNPVQAELQMLWHLMYLNGSYVANDPEANFDGVRVD 465
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 56 IDVTLNDNIKIV-DYIDGS-DIDGS-----DGWFTYGNPNENY-----TI 94
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 466 AVDNVANADLQIADSVLKAYGVDSKSKNAINHLSTLEAM--SDNDPOYKDKTGGAQLPI 523
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 95 PNSQSYSL-----LKSNSQITQIKRYVSCNSTRLKRFKSPSAKYTTTSGKVISITON 147
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 524 DNYLRISLVALTRPLEKDSNKNRISGLPEVITNSLNKRSLEKGNSEPMANYITIRAH 583
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 148 SINSSRVVINA-----DATNPTDELR-----TKETRENOSYT----- 183
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 584 DSEVQVIAKIIVIAQINPKTDGLFTLDELKQAFKLYNEDMRQAKKKTQNSIPTAYALM 643
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 184 -SHKSTNSLY-----VHTWTIPR-----SLKQNMWRMEDY 213
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 644 LSNKDSITRLLYGDMYSDGQYVATKSPYYDAIDTLKARIKXAAAGQDMKITTYVEGDKS 703
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 214 NNGWTWA-----QSCYVTGA-----DGSESTRWLAAGSIPPPGNVDGLWL---DNDIALSG 263
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Db 704 HMDDDYGVLTSVYRGANENATDQSEATK--TQGAIVTISNNPSLKLNQNDKVIYVMG 761

QY 264 MARKS 268  
Db 762 AAHKN 766

## RESULT 13

US-09-210-361-6  
Sequence 6, Application US/09210361

Patent No. 6284479

GENERAL INFORMATION:

APPLICANT: Nichols, Scott E.

TITLE OF INVENTION: Substitutes for Modified Starches and

TITLE OF INVENTION: Latexes in Paper Manufacture

FILE REFERENCE: 0357CR

CURRENT APPLICATION NUMBER: US/09/210,361

EARLIER FILING DATE: 1998-12-11

EARLIER FILING DATE: 09/007,999

EARLIER FILING DATE: 1998-01-16

EARLIER FILING DATE: 08/478,704

EARLIER FILING DATE: 1995-06-07

EARLIER FILING DATE: 09/009,620

EARLIER FILING DATE: 1998-01-20

EARLIER FILING DATE: 08/485,243

EARLIER FILING DATE: 1995-06-07

EARLIER FILING DATE: 09/008,172

EARLIER FILING DATE: 1998-01-16

EARLIER FILING DATE: 08/482,711

NUMBER OF SEQ ID NOS: 6

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 6

LENGTH: 1430

TYPE: PRT

ORGANISM: streptococcus mutans

US-09-210-361-6

Query Match 5.9%; Score 105; DB 2; Length 1430;

Best Local Similarity 22.5%; Pred. No. 1.1;

Matches 82; Conservative 40; Mismatches 133; Indels 110; Gaps 19;

QY 5 EDMSDWTGVSTFNVILETGLDNCN--IYANGLMNIGVIIN--ITPTDDEGNF---VD 55

Db 411 EDNS-----GGYDFLLANDIDNSNPVQAEOQMLHLYMNGSIVANDPEANFDGVRVD 465

QY 56 IDVTYLDNINIKIV-DYIDGS-DIDGS-----DGMFYGNPNREYN-----TI 94

Db 466 AVDNVNADLLQIASDYLAHAGVDSKKNALNHLSILEAM--SDNDPQYNKDTGAQLPI 523

QY 95 PMSQSYSL-----LKSNSQITQIRYVSCNSTRSLRTKSFSAKYTTTSGKVISITON 147

Db 524 DNKRLSLVLAITRPLEKDSNKNRISGLEPVTNSLNRSAGKXSERBANYIFIRAH 583

QY 148 SINSSRVVINA-----DATNFTDDELRT-----TKETRFENQSYT----- 183

Db 544 DSEVQTVIAKTIKAQINPKTDGLFTLDELKQAFKIYNEDEKQAKKKTOSNIPTAYALM 643

QY 184 -SHKSTNSLY-----VHTWTIIPR-----SLKQNMREWDY 213

Db 644 LSNKDSITRLYYGDMVSDDGQYMATKSPYYDAIDTLKARKIKYAAGGQDMKITVEGDKS 703

QY 214 NNGWTMA---QSCYKTTGA---DGSSESTRWLAAGSIFPPGNVGLML---DNDIALSG 263

Db 704 HMDDDYGVLTSVYRGANENATDQSEATK--TQGAIVTISNNPSLKLNQNDKVIYVMG 761

QY 264 MARKS 268

Db 762 AAHKN 766

RESULT 14

US-09-740-274-6  
Sequence 6, Application US/09740274

Patent No. 6465203

GENERAL INFORMATION:

APPLICANT: Nichols, Scott E.

TITLE OF INVENTION: Glucan-containing Compositions and Paper

FILE REFERENCE: 0357CRD

CURRENT APPLICATION NUMBER: US/09/740,274

EARLIER FILING DATE: 2000-12-19

EARLIER FILING DATE: 09/210,361

EARLIER FILING DATE: 1998-12-11

EARLIER FILING DATE: 09/007,999

EARLIER FILING DATE: 1998-01-16

EARLIER FILING DATE: 08/478,704

EARLIER FILING DATE: 1995-06-07

EARLIER FILING DATE: 09/009,620

EARLIER FILING DATE: 1998-01-20

EARLIER FILING DATE: 08/485,243

EARLIER FILING DATE: 1995-06-07

EARLIER FILING DATE: 09/008,172

EARLIER FILING DATE: 1998-01-16

EARLIER FILING DATE: 08/482,711

NUMBER OF SEQ ID NOS: 6

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 6

LENGTH: 1430

TYPE: PRT

ORGANISM: streptococcus mutans

US-09-740-274-6

Query Match 5.9%; Score 105; DB 2; Length 1430;

Best Local Similarity 22.5%; Pred. No. 1.1;

Matches 82; Conservative 40; Mismatches 133; Indels 110; Gaps 19;

QY 5 EDMSDWTGVSTFNVILETGLDNCN--IYANGLMNIGVIIN--ITPTDDEGNF---VD 55

Db 411 EDNS-----GGYDFLLANDIDNSNPVQAEOQMLHLYMNGSIVANDPEANFDGVRVD 465

QY 56 IDVTYLDNINIKIV-DYIDGS-DIDGS-----DGMFYGNPNREYN-----TI 94

Db 466 AVDNVNADLLQIASDYLAHAGVDSKKNALNHLSILEAM--SDNDPQYNKDTGAQLPI 523

QY 95 PMSQSYSL-----LKSNSQITQIRYVSCNSTRSLRTKSFSAKYTTTSGKVISITON 147

Db 524 DNKRLSLVLAITRPLEKDSNKNRISGLEPVTNSLNRSAGKXSERBANYIFIRAH 583

QY 148 SINSSRVVINA-----DATNFTDDELRT-----TKETRFENQSYT----- 183

Db 584 DSEVQTVIAKTIKAQINPKTDGLFTLDELKQAFKIYNEDEKQAKKKTOSNIPTAYALM 643

QY 184 -SHKSTNSLY-----VHTWTIIPR-----SLKQNMREWDY 213

Db 644 LSNKDSITRLYYGDMVSDDGQYMATKSPYYDAIDTLKARKIKYAAGGQDMKITVEGDKS 703

QY 214 NNGWTMA---QSCYKTTGA---DGSSESTRWLAAGSIFPPGNVGLML---DNDIALSG 263

Db 704 HMDDDYGVLTSVYRGANENATDQSEATK--TQGAIVTISNNPSLKLNQNDKVIYVMG 761

QY 264 MARKS 268

Db 762 AAHKN 766

## RESULT 15

US-09-538-092-330  
Sequence 330, Application US/09538092

Patent No. 6753314

GENERAL INFORMATION:

APPLICANT: Glot, Ioic

APPLICANT: Mansfield, Traci A.

TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same

FILE REFERENCE: 15966-542





THIS PAGE LEFT BLANK

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: January 30, 2006, 09:43:19 ; Search time 20.8022 Seconds

(without alignments)  
6708.668 Million cell updates/sec

Title: US-09-889-874a-22

Perfect score: 1784

Sequence: 1 FTIRSDMSMDWTGVSTFNV.....ILFHAGYNSDPYLSSTIVY 334

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417629326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Published Applications AA Main:  
1: /cgn2\_6/ptodaca/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodaca/1/pubpaa/US08\_PUBCOMB.pep:\*  
3: /cgn2\_6/ptodaca/1/pubpaa/US09\_PUBCOMB.pep:\*  
4: /cgn2\_6/ptodaca/1/pubpaa/US10A\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodaca/1/pubpaa/US10B\_PUBCOMB.pep:\*  
6: /cgn2\_6/ptodaca/1/pubpaa/US11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	126	7.1	892	US-10-282-122A-70481	Sequence 70481, A
2	126	7.1	892	US-10-661-809-21	Sequence 21, App1
3	126	7.1	930	US-10-615-383-10	Sequence 10, App1
4	126	7.1	930	US-10-680-184-10	Sequence 10, App1
5	126	7.1	930	US-10-689-082-10	Sequence 10, App1
6	126	7.1	930	US-10-724-972A-5309	Sequence 5309, App1
7	124	7.0	1092	US-10-378-674-8	Sequence 8, App1
8	124	7.0	1092	US-10-806-288-15	Sequence 15, App1
9	118.5	6.6	670	US-10-282-122A-70444	Sequence 70444, A
10	118	6.6	6761	US-10-732-823-15035	Sequence 15035, A
11	115.5	6.5	1411	US-10-080-505-17	Sequence 17, App1
12	115.5	6.5	1411	US-10-687-046-17	Sequence 17, App1
13	114	6.4	360	US-10-617-320-5178	Sequence 5178, App1
14	112.5	6.3	1436	US-10-080-505-13	Sequence 13, App1
15	112.5	6.3	1436	US-10-687-046-13	Sequence 13, App1
16	111	6.2	369	US-10-472-928-1818	Sequence 1818, App1
17	109.5	6.1	285	US-10-452-024-65	Sequence 65, App1
18	109	6.1	807	US-09-820-843A-108	Sequence 108, App1
19	108	6.1	626	US-10-739-930-6661	Sequence 6661, App1
20	108	6.1	978	US-10-282-122A-75356	Sequence 75356, A
21	107.5	6.0	324	US-10-994-726-418	Sequence 418, App1
22	107.5	6.0	324	US-10-994-726-417	Sequence 417, App1
23	107.5	6.0	623	US-10-452-024-73	Sequence 73, App1
24	107.5	6.0	667	US-10-425-115-292261	Sequence 292261, App1
25	107.5	6.0	753	US-10-425-114-54358	Sequence 54358, A
26	107.5	6.0	1038	US-10-282-122A-43827	Sequence 43827, A
27	107	6.0	1449	US-10-282-122A-65711	Sequence 65711, A

28	106.5	6.0	623	US-10-452-024-72	Sequence 72, App1
29	106.5	6.0	738	US-10-425-115-332566	Sequence 332566, A
30	106.5	6.0	619	US-10-425-114-56258	Sequence 56258, A
31	105.5	5.9	991	US-10-831-070-54	Sequence 54, App1
32	105	5.9	507	US-10-712-533A-10	Sequence 10, App1
33	105	5.9	1430	US-09-740-274-6	Sequence 6, App1
34	105	5.9	1430	US-10-383-930-36	Sequence 36, App1
35	105	5.9	1430	US-10-797-821-36	Sequence 36, App1
36	104.5	5.9	643	US-10-369-493-15821	Sequence 15821, A
37	104.5	5.9	643	US-10-369-493-15453	Sequence 15453, A
38	104.5	5.9	639	US-10-369-493-15453	Sequence 13, App1
39	103.5	5.8	582	US-10-806-288-13	Sequence 11, App1
40	103.5	5.8	593	US-10-806-288-11	Sequence 11, App1
41	103.5	5.8	1031	US-10-369-493-22004	Sequence 22004, A
42	103.5	5.8	1435	US-10-484-703-32	Sequence 32, App1
43	103.5	5.8	1481	US-10-472-928-1236	Sequence 1236, App1
44	103	5.8	1395	US-10-080-505-7	Sequence 7, App1
45	103	5.8	1395	US-10-687-046-7	Sequence 7, App1

## ALIGNMENTS

### RESULT 1

US-10-282-122A-70481  
Sequence 70481, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remainder of Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 70481  
LENGTH: 892  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-10-282-122A-70481



```

; Sequence 10, Application US/10690184
; Publication No. US20040141997A1
; GENERAL INFORMATION:
; APPLICANT: POSTER, Timothy
; TITLE OF INVENTION: METHODS FOR TREATING OR PREVENTING INFECTIONS FROM COAGULASE-
; FILE REFERENCE: P06335US05/BAS
; CURRENT APPLICATION NUMBER: US/10/690,184
; CURRENT FILING DATE: 2003-10-21
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: 60/098,443
; PRIOR FILING DATE: 1998-08-31
; PRIOR APPLICATION NUMBER: 60/117,119
; PRIOR FILING DATE: 1999-01-25
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 930
; TYPE: PR
; ORGANISM: Staphylococcus epidermidis
US-10-690-184-10

```

```

Query Match      7.1%; Score 126; DB 4; Length 930;
Best Local Similarity 21.0%; Pred. No. 0.11;
Matches 78; Conservative 51; Mismatches 141; Indels 102; Gaps 16;

```

```

QY 9 SDMTGVSTFNVILETGLDNCNIYANGIANGMIG--VIINPTDDEGNFVIDDVT----- 60
DB 436 SMETNIDTGNHYEQ-----TIYNPLRSAGKNTNINISGNDGSGTI-IDDSIIIKYVK 489
QY 61 -----LNDNIKIVDYIDSGDIDGSDGWFYTGPNENYNT-----IPNS 97
DB 490 VGDNQMLPDSNRILYDSEYEDVT--NDVYAQLGNNNDVNINFGNIDSPYIIKVISKYDPNK 548
QY 98 QSYSLKSENSQITQIKRYVSCNTRSLRTKFSKAV--TTTSG-----KVISI-- 144
DB 549 DDYTTIQOTVTWQTINEX-----TGEFRPASYDNTIAFSTSSGQGGDLPEPKTYKIGD 603
QY 145 -----TONSINSRVAIVNAIDATNFTDDELRTKTRPENOSYSHKSTNSL-Y 193
DB 604 YWEDVDKDGICQNTNDEKPLSNVLVLTYPDG--TSKSVRTDEGKIQFDGLKXGLTY 660
QY 194 VHTWTIPRSLKLNWMEDEY-----NNGWTWASQCYRTGADGSGEST 236
DB 661 KITFERP-----EGYPTLKSGTNPALDSEGNVWVTINGQDDMTIDSGFYQTP 710
QY 237 RMLAASIFPPGNYDGLMDNDIALSGMAHKSYNVDTGINQLSFTRIIGKGFWMYVNIISG 296
DB 711 KYSLGNYVWYDTRKDGICQDDEKIGISGV--KVLKDENGNIISTTTTDENG--KIQFDN 765
QY 297 LDRGHAVIITDQ 308
DB 766 LNSGNVIVHFDK 777

```

```

RESULT 5
US-10-689-082-10
; Sequence 10, Application US/10689082
; Publication No. US20040142348A1
; GENERAL INFORMATION:
; APPLICANT: POSTER, Timothy
; TITLE OF INVENTION: POLYPEPTIDES AND POLYNUCLEOTIDES FROM COAGULASE-NEGATIVE STAPHYLO-
; FILE REFERENCE: P06335US04/BAS
; CURRENT APPLICATION NUMBER: US/10/689,082
; CURRENT FILING DATE: 2003-10-21
; PRIOR APPLICATION NUMBER: 09/386,962
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: 60/098,443
; PRIOR FILING DATE: 1998-08-31
; PRIOR APPLICATION NUMBER: 60/117,119
; PRIOR FILING DATE: 1999-01-25
; NUMBER OF SEQ ID NOS: 39

```

```

; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 930
; TYPE: PR
; ORGANISM: Staphylococcus epidermidis
US-10-689-082-10

```

```

Query Match      7.1%; Score 126; DB 4; Length 930;
Best Local Similarity 21.0%; Pred. No. 0.11;
Matches 78; Conservative 51; Mismatches 141; Indels 102; Gaps 16;

```

```

QY 9 SDMTGVSTFNVILETGLDNCNIYANGIANGMIG--VIINPTDDEGNFVIDDVT----- 60
DB 436 SMETNIDTGNHYEQ-----TIYNPLRSAGKNTNINISGNDGSGTI-IDDSIIIKYVK 489
QY 61 -----LNDNIKIVDYIDSGDIDGSDGWFYTGPNENYNT-----IPNS 97
DB 490 VGDNQMLPDSNRILYDSEYEDVT--NDVYAQLGNNNDVNINFGNIDSPYIIKVISKYDPNK 548
QY 98 QSYSLKSENSQITQIKRYVSCNTRSLRTKFSKAV--TTTSG-----KVISI-- 144
DB 549 DDYTTIQOTVTWQTINEX-----TGEFRPASYDNTIAFSTSSGQGGDLPEPKTYKIGD 603
QY 145 -----TONSINSRVAIVNAIDATNFTDDELRTKTRPENOSYSHKSTNSL-Y 193
DB 604 YWEDVDKDGICQNTNDEKPLSNVLVLTYPDG--TSKSVRTDEGKIQFDGLKXGLTY 660
QY 194 VHTWTIPRSLKLNWMEDEY-----NNGWTWASQCYRTGADGSGEST 236
DB 661 KITFERP-----EGYPTLKSGTNPALDSEGNVWVTINGQDDMTIDSGFYQTP 710
QY 237 RMLAASIFPPGNYDGLMDNDIALSGMAHKSYNVDTGINQLSFTRIIGKGFWMYVNIISG 296
DB 711 KYSLGNYVWYDTRKDGICQDDEKIGISGV--KVLKDENGNIISTTTTDENG--KIQFDN 765
QY 297 LDRGHAVIITDQ 308
DB 766 LNSGNVIVHFDK 777

```

```

RESULT 6
US-10-724-972A-5309
; Sequence 5309, Application US/10724972A
; Publication No. US20040147734A1
; GENERAL INFORMATION:
; APPLICANT: Doucelte-Stamm, Lynn
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: PAT03-16
; CURRENT APPLICATION NUMBER: US/10/724,972A
; CURRENT FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: 09/450,969
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 7544
; SEQ ID NO 5309
; LENGTH: 930
; TYPE: PR
; ORGANISM: S. epidermidis
US-10-724-972A-5309

```

```

Query Match      7.1%; Score 126; DB 4; Length 930;
Best Local Similarity 21.0%; Pred. No. 0.11;
Matches 78; Conservative 51; Mismatches 141; Indels 102; Gaps 16;

```

```

QY 9 SDMTGVSTFNVILETGLDNCNIYANGIANGMIG--VIINPTDDEGNFVIDDVT----- 60
DB 436 SMETNIDTGNHYEQ-----TIYNPLRSAGKNTNINISGNDGSGTI-IDDSIIIKYVK 489
QY 61 -----LNDNIKIVDYIDSGDIDGSDGWFYTGPNENYNT-----IPNS 97
DB 490 VGDNQMLPDSNRILYDSEYEDVT--NDVYAQLGNNNDVNINFGNIDSPYIIKVISKYDPNK 548
QY 98 QSYSLKSENSQITQIKRYVSCNTRSLRTKFSKAV--TTTSG-----KVISI-- 144
DB 549 DDYTTIQOTVTWQTINEX-----TGEFRPASYDNTIAFSTSSGQGGDLPEPKTYKIGD 603
QY 145 -----TONSINSRVAIVNAIDATNFTDDELRTKTRPENOSYSHKSTNSL-Y 193
DB 604 YWEDVDKDGICQNTNDEKPLSNVLVLTYPDG--TSKSVRTDEGKIQFDGLKXGLTY 660
QY 194 VHTWTIPRSLKLNWMEDEY-----NNGWTWASQCYRTGADGSGEST 236
DB 661 KITFERP-----EGYPTLKSGTNPALDSEGNVWVTINGQDDMTIDSGFYQTP 710
QY 237 RMLAASIFPPGNYDGLMDNDIALSGMAHKSYNVDTGINQLSFTRIIGKGFWMYVNIISG 296
DB 711 KYSLGNYVWYDTRKDGICQDDEKIGISGV--KVLKDENGNIISTTTTDENG--KIQFDN 765
QY 297 LDRGHAVIITDQ 308
DB 766 LNSGNVIVHFDK 777

```

Db 436 SMFTNIDTKNHTVEQ-----TIYINPLRYSAKETNVNISGNDEGSTI-IDDSTIIKVKY 489  
QY 61 -----LNDIKIKVDYIDSGDIDSGDMFTYGNPNENT-----IPNS 97  
Db 490 VGDNONLPDSNRKYDYSEYEDVT-NDYAOQLGNNDVNINFGNIDSPYIIKVISKIDPNK 548  
QY 98 QSYSLKSENSQITQIKRYVSCSNTSRLRTKSPSAKV--TTTSG-----KVISI-- 144  
Db 549 DDYTTIQOTYVMQTTINEY-----TGEFRITASYDNTIAFSTSSGQGGDLPEPKTYKIGD 603  
QY 145 -----TONSINSRVRVINAIDATNFTDDELRTTKETREPNOSYTSKSKSTNSL-Y 193  
Db 604 YWEDVDKDGIOQNTNDEKPLSNVLVLTYPDG---TSKSVRTDEBKCYQFDGLKNGLTY 660  
QY 194 VHTWTIPRSLKLGQWREDEY-----NNGMTWAQSCYYTKGADGSGEST 236  
Db 661 KITFETP-----EGTTPILKHSCTNPALDSEGNISVWVTINGDDMTIDSGFYQTP 710  
QY 237 RWLAAGSIFPPGNYVDGLMNDIALSGMAHKSYNVDTGINQLSFTRIIGKGFPMVYNISG 296  
Db 711 KYSLGNYVWYDTNKDGIQGDDEKIGISGV--KYTLKDENGNIISTTTDENG---KYQFDN 765  
QY 297 LDRGHAVIITDQ 308  
Db 766 LNSGNYIVHFDK 777

## RESULT 7

US-10-378-674-8  
; Sequence 8, Application US/10378674  
; Publication No. US20040006209A1  
; GENERAL INFORMATION:  
; APPLICANT: PATTI, Joseph M.  
; TITLE OF INVENTION: MONOCLONAL AND POLYCLONAL ANTIBODIES RECOGNIZING COAGULASE-NEGATI  
; TITLE OF INVENTION: STAPHYLOCOCCAL PROTEINS  
; FILE REFERENCE: P07556U501/BAS  
; CURRENT APPLICATION NUMBER: US/10/378,674  
; CURRENT FILING DATE: 2003-03-05  
; PRIOR APPLICATION NUMBER: 60/361,324  
; PRIOR FILING DATE: 2002-03-05  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 8  
; LENGTH: 1092  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-10-378-674-8

Query Match 7.0%; Score 124; DB 4; Length 1092;

Best local similarity 21.0%; Pred. No. 0.2;

Matches 78; Conservative 50; Mismatches 142; Indels 102; Gaps 16;

QY 9 SDMTGVSTFNVILETGLDNCNIYANGLMIG--VIINITPDDSGNFVDIDVT-----60  
Db 439 SMFTNIDTKNHTVEQ-----TIYINPLRYSAKETNVNISGNDEGSTI-IDDSTIIKVKY 492  
QY 61 -----LNDIKIKVDYIDSGDIDSGDMFTYGNPNENT-----IPNS 97  
Db 493 VGDNONLPDSNRKYDYSEYEDVT-NDYAOQLGNNDVNINFGNIDSPYIIKVISKIDPNK 551  
QY 98 QSYSLKSENSQITQIKRYVSCSNTSRLRTKSPSAKV--TTTSG-----KVISI-- 144  
Db 552 DDYTTIQOTYVMQTTINEY-----TGEFRITASYDNTIAFSTSSGQGGDLPEPKTYKIGD 606  
QY 145 -----TONSINSRVRVINAIDATNFTDDELRTTKETREPNOSYTSKSKSTNSL-Y 193  
Db 607 YWEDVDKDGIOQNTNDEKPLSNVLVLTYPDG---TSKSVRTDEBKCYQFDGLKNGLTY 663  
QY 194 VHTWTIPRSLKLGQWREDEY-----NNGMTWAQSCYYTKGADGSGEST 236  
Db 664 KITFETP-----EGTTPILKHSCTNPALDSEGNISVWVTINGDDMTIDSGFYQTP 713  
QY 237 RWLAAGSIFPPGNYVDGLMNDIALSGMAHKSYNVDTGINQLSFTRIIGKGFPMVYNISG 296

Db 714 KYSLGNYVWYDTNKDGIQGDDEKIGISGV--KYTLKDENGNIISTTTDENG---KYQFDN 768  
QY 297 LDRGHAVIITDQ 308  
Db 769 LNSGNYIVHFDK 780

## RESULT 8

US-10-806-288-15

; Sequence 15, Application US/10806288

; Publication No. US20040209326A1

; GENERAL INFORMATION:

; APPLICANT: Guss, Bengt

; APPLICANT: Nilsson, Martin

; APPLICANT: Frykberg, Lars

; APPLICANT: Flock, Jan-Ingemar

; APPLICANT: Lindberg, Martin

; TITLE OF INVENTION: Fibrinogen Binding Protein Originating from

; FILE REFERENCE: Guss 09/147405

; CURRENT APPLICATION NUMBER: US/10/806,288

; CURRENT FILING DATE: 2004-03-23

; PRIOR APPLICATION NUMBER: US/09/147,405

; PRIOR FILING DATE: 1999-04-11

; PRIOR APPLICATION NUMBER: PCT/SE97/10191

; PRIOR FILING DATE: 1997-06-18

; PRIOR APPLICATION NUMBER: SE 9602496-3

; PRIOR FILING DATE: 1996-06-20

; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 15

; LENGTH: 1092

; TYPE: PRT

; ORGANISM: Staphylococcus epidermidis

US-10-806-288-15

Query Match 7.0%; Score 124; DB 4; Length 1092;

Best local similarity 21.0%; Pred. No. 0.2;

Matches 78; Conservative 50; Mismatches 142; Indels 102; Gaps 16;

QY 9 SDMTGVSTFNVILETGLDNCNIYANGLMIG--VIINITPDDSGNFVDIDVT-----60  
Db 439 SMFTNIDTKNHTVEQ-----TIYINPLRYSAKETNVNISGNDEGSTI-IDDSTIIKVKY 492  
QY 61 -----LNDIKIKVDYIDSGDIDSGDMFTYGNPNENT-----IPNS 97  
Db 493 VGDNONLPDSNRKYDYSEYEDVT-NDYAOQLGNNDVNINFGNIDSPYIIKVISKIDPNK 551  
QY 98 QSYSLKSENSQITQIKRYVSCSNTSRLRTKSPSAKV--TTTSG-----KVISI-- 144  
Db 552 DDYTTIQOTYVMQTTINEY-----TGEFRITASYDNTIAFSTSSGQGGDLPEPKTYKIGD 606  
QY 145 -----TONSINSRVRVINAIDATNFTDDELRTTKETREPNOSYTSKSKSTNSL-Y 193  
Db 607 YWEDVDKDGIOQNTNDEKPLSNVLVLTYPDG---TSKSVRTDEBKCYQFDGLKNGLTY 663  
QY 194 VHTWTIPRSLKLGQWREDEY-----NNGMTWAQSCYYTKGADGSGEST 236  
Db 664 KITFETP-----EGTTPILKHSCTNPALDSEGNISVWVTINGDDMTIDSGFYQTP 713  
QY 237 RWLAAGSIFPPGNYVDGLMNDIALSGMAHKSYNVDTGINQLSFTRIIGKGFPMVYNISG 296  
Db 714 KYSLGNYVWYDTNKDGIQGDDEKIGISGV--KYTLKDENGNIISTTTDENG---KYQFDN 768  
QY 297 LDRGHAVIITDQ 308  
Db 769 LNSGNYIVHFDK 780

## RESULT 9

US-10-282-122A-70444

; Sequence 70444, Application US/10282122A

```

Db      282 DTMKDQIGDPDEKGISGV--KYTLTKDNNGNIIISTTTDENG---KYQFDNLNSGNYLVHF 336
QY      307 DQ 308
       |||
Db      337 DK 338

RESULT 10
US-10-732-923-15035
; Sequence 15035, Application US/10732923
; Publication No. US20050108791A1
GENERAL INFORMATION:
APPLICANT: Edgerton, Michael D
TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
FILE REFERENCE: 38-15(52796)C
CURRENT APPLICATION NUMBER: US/10/732,923
CURRENT FILING DATE: 2003-12-10
PRIOR APPLICATION NUMBER: 10/310,154
PRIOR FILING DATE: 2002-12-04
NUMBER OF SEQ ID NOS: 24149
SEQ ID NO 15035
LENGTH: 6761
TYPE: PRF
ORGANISM: Plasmodium falciparum
US-10-732-923-15035

Query Match          6.6%; Score 118; DB 5; Length 6761;
Best Local Similarity 21.9%; Pred. No. 7.4;
Matches   80; Conservative 49; Mismatches 123; Indels 114; Gaps 18;

QY      17 ENVILETGLDNCNIYANGLMIGVIINIPTPDEGNFVDIDV--TLNDIKIVDI DG 74
       |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      1113 FNCIIESEFPENYDINVEG-----EDGNFYVPKNKLIRKVTITNETDSS 1158

QY      75 DI-----DSGDGWFTYGPNRBNYNTIPNSQSYSLKSNSQITOIK----RYVGSNTSRL 125
       |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      1159 DMVIELKDBERKGFTCIDYRVSPF-----LKKONKKGAQVKGIIRKYS-----P 1202

QY      126 RTKSF-----SAKVTTTSQKVISITONSINSRHVINAIADATNFDTDELRTTKTRF-- 177
       |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      1203 KYIQKFPIHLELVNKLYGTSD--LSKGNNNNNNNM-----KENGVK 1242

QY      178 ---ENOSYTHSKSTNSLYVHTW-TTPRSLLKLNWRMEDYRNNGWTMAQSCTYTGDAGGS 233
       |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      1243 CGHSYSRASCKSISSISYINKKYAVELNEKITIYIIDDDNNNNIN--SCCTYKNNLSNEN 1300

QY      234 ESTRWLAAGSIFPPRGVTDGLMDNDIALSGMAHKSVYVDG---INLSFTRIILKGFWS 290
       |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      1301 N---LCADRVLDGDYN--MLENDV-----DEMNVDONOKKRNDLVSDVGKWSPY 1348

QY      291 VVNISGLDR-----GHAVIIDIQGN-----KYLIFPHAGYNSDP 326
       |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      1349 FYNIKLIANKPEPRPILEKCDKGGEMTDHVANTLKNEICISCRILLFMGHGHDYSC 1408

QY      327 YLSSSI 332
       |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      1409 SLSSAI 1414

RESULT 11
US-10-080-505-17
; Sequence 17, Application US/10080505
; Publication No. US20030073166A1
GENERAL INFORMATION:
APPLICANT: St. Gene, Joseph W.
TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
FILE REFERENCE: A-59941-1/RET/DCF/DHR
CURRENT APPLICATION NUMBER: US/10/080,505
CURRENT FILING DATE: 2002-02-22
PRIOR APPLICATION NUMBER: US 08/226,791
PRIOR FILING DATE: 1994-10-25
PRIOR APPLICATION NUMBER: US 09/839,996
PRIOR FILING DATE: 2001-04-20

```

NUMBER OF SEQ ID NOS: 58  
 SOFTWARE: PatentIn version 3.1  
 SEQ ID NO 17  
 LENGTH: 1411  
 TYPE: PRT  
 ORGANISM: Haemophilus influenzae  
 US-10-080-505-17

Query Match 6.5%; Score 115.5; DB 4; Length 1411;  
 Best Local Similarity 20.8%; Pred. No. 1.5;  
 Matches 85; Conservative 48; Mismatches 131; Indels 145; Gaps 21;

QY 4 REDSMSPMTGVS-----TFNVILETGLDNCNIYANGLMIGVINIT-----45  
 DB 554 RDIAPNMGFDKDTKMTGRNLAVTNPL--NKNHFLISGDTLKG---NITDGGTLV 607  
 QY 46 ----PTDEGNFVD-----IDVTLNDNIKIYDY-IDG-----SDIDG 78  
 DB 608 FSGRPPTPAVNHNLRLNELGRPKGEVVIDDWMIRTFKAENFQIKGGSTVSRNVSSIEG 667  
 QY 79 SDGPFYTGPNV-EYNTIPNSQSYSLKSENSQITQIKRYVSCNMTSLRTKSPSAKYTTT 137  
 DB 668 N-WTISNNANATGVPVNOQNTICTRSDWTGLT-----TCKTVNLTDKKVIDSIPTTQ 719  
 QY 138 SGKVISITONS-----INSSRVIN-----AIDATN- 163  
 DB 720 INGSINLTNATVINHGLAKNGVTLINHSQFTLSNNATQTNIGLSNNANATVDANL 779  
 QY 164 ----FTDDELRTKTRTFENQSYTSKSSSTNSLYHTWTIPSLKQNNRMEDYNNGWT 218  
 DB 780 NGNVHLTDSAQFSLKNSHFHQ-IOGDKDTVTYLENATWTPSDTTLQNL--TLNNSTV 835  
 QY 219 WAOSCY-----YKTGADGSESTRW---LAASIFPPGNYDGL-----WLDN 257  
 DB 836 TLNSAVSASSNNAPRRHRSLETETPTSEHRFNLTIVNGKLSGGCTFTQFTSSILFGYKSD 895  
 QY 258 DIALSGMAHKSYNV---DTG-----INQLSFTRIIGKFSWVYNISGLD 298  
 DB 896 KIKLSNDAEGDYTLAVRDTGKEPVTLEQLTL-----IEGLD 931

RESULT 12  
 US-10-687-046-17  
 Sequence 17, Application US/10687046  
 Publication No. US20040157241A1  
 GENERAL INFORMATION:  
 APPLICANT: St. Gene, Joseph W.  
 TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS  
 FILE REFERENCE: A-59941-1/RT/DCF/DRR  
 CURRENT APPLICATION NUMBER: US/10/687,046  
 CURRENT FILING DATE: 2003-10-15  
 PRIOR APPLICATION NUMBER: US/10/080,505  
 PRIOR FILING DATE: 2002-02-22  
 PRIOR APPLICATION NUMBER: US 08/296,791  
 PRIOR FILING DATE: 1994-10-25  
 PRIOR APPLICATION NUMBER: US 09/839,996  
 PRIOR FILING DATE: 2001-04-20  
 NUMBER OF SEQ ID NOS: 58  
 SOFTWARE: PatentIn version 3.1  
 SEQ ID NO 17  
 LENGTH: 1411  
 TYPE: PRT  
 ORGANISM: Haemophilus influenzae  
 US-10-687-046-17

Query Match 6.5%; Score 115.5; DB 4; Length 1411;  
 Best Local Similarity 20.8%; Pred. No. 1.5;  
 Matches 85; Conservative 48; Mismatches 131; Indels 145; Gaps 21;

QY 4 REDSMSPMTGVS-----TFNVILETGLDNCNIYANGLMIGVINIT-----45  
 DB 554 RDIAPNMGFDKDTKMTGRNLAVTNPL--NKNHFLISGDTLKG---NITDGGTLV 607

QY 46 ----PTDEGNFVD-----IDVTLNDNIKIYDY-IDG-----SDIDG 78  
 DB 608 FSGRPPTPAVNHNLRLNELGRPKGEVVIDDWMIRTFKAENFQIKGGSTVSRNVSSIEG 667  
 QY 79 SDGPFYTGPNV-EYNTIPNSQSYSLKSENSQITQIKRYVSCNMTSLRTKSPSAKYTTT 137  
 DB 668 N-WTISNNANATGVPVNOQNTICTRSDWTGLT-----TCKTVNLTDKKVIDSIPTTQ 719  
 QY 138 SGKVISITONS-----INSSRVIN-----AIDATN- 163  
 DB 720 INGSINLTNATVINHGLAKNGVTLINHSQFTLSNNATQTNIGLSNNANATVDANL 779  
 QY 164 ----FTDDELRTKTRTFENQSYTSKSSSTNSLYHTWTIPSLKQNNRMEDYNNGWT 218  
 DB 780 NGNVHLTDSAQFSLKNSHFHQ-IOGDKDTVTYLENATWTPSDTTLQNL--TLNNSTV 835  
 QY 219 WAOSCY-----YKTGADGSESTRW---LAASIFPPGNYDGL-----WLDN 257  
 DB 836 TLNSAVSASSNNAPRRHRSLETETPTSEHRFNLTIVNGKLSGGCTFTQFTSSILFGYKSD 895  
 QY 258 DIALSGMAHKSYNV---DTG-----INQLSFTRIIGKFSWVYNISGLD 298  
 DB 896 KIKLSNDAEGDYTLAVRDTGKEPVTLEQLTL-----IEGLD 931

RESULT 13  
 US-10-617-320-5178  
 Sequence 5178, Application US/10617320  
 Publication No. US20050136404A1  
 GENERAL INFORMATION:  
 APPLICANT: Lynn A Doucette-Stamm and David Bush  
 TITLE OF INVENTION: NOCULEIC ACID AND AMINO ACID  
 SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGNOSIS  
 THERAPEUTICS

NUMBER OF SEQUENCES: 5206  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
 STREET: 100 Beaver Street  
 CITY: Waltham  
 STATE: Massachusetts  
 COUNTRY: USA  
 ZIP: 02354

COMPUTER READABLE FORM:  
 MEDIUM TYPE: CD-ROM ISO9660  
 COMPUTER: <Unknown>  
 OPERATING SYSTEM: <Unknown>  
 SOFTWARE: <Unknown>

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/10/617,320  
 FILING DATE: 10-Jul-2003  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/09/107,433  
 FILING DATE: 30-Jun-1998

APPLICATION NUMBER: 60/085131  
 FILING DATE: May 12, 1998  
 APPLICATION NUMBER: 60/051553  
 FILING DATE: July 2, 1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Ariniello, Pamela Deneke  
 REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-011  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (781)893-5007  
 TELEFAX: (781)893-8277  
 INFORMATION FOR SEQ ID NO: 5178:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 360 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULAR TYPE: protein  
 HYPOTHETICAL: YES  
 ORIGINAL SOURCE:  
 ORGANISM: Streptococcus pneumoniae



FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...360  
SEQUENCE DESCRIPTION: SEQ ID NO: 5178  
US-10-617-320-5178

Query Match 6.4%; Score 114; DB 5; Length 360;

Best Local Similarity 25.6%; Pred. No. 0.33; Mismatches 73; Indels 28; Gaps 8;

Matches 46; Conservative 33; Mismatches 73; Indels 28; Gaps 8;

31 YANGLMGVIIN--ITPTDEGNFVIDDVTINDNIKIYIDG-----SDIDG 78  
104 YKNAHQALSTLNLGALPIINENDSVIVDEKVGDNPTLSAQVAAVQADLLVLTVDVG 163  
79 SDGMPYGNENEVYITNSQSYSLKSENGQITQIKRYVCSN-TSLRTKSPSAKYTTT 137  
164 ----LYTGNP--NSDPRARLERIETIRREIIDMAGAGSSNGTGMLTKIKATYATE 216  
138 SGKVISITONSINSSRVVINAIDATNFTDDE--LRTTKETRFENQSYSHKSTNSLYV 194  
217 SGVPVYIC-SLSKSD---SMIEAETEDGSGYVVAQEKLRTOXOMLAFYAGSQSISIW 271

RESULT 14

US-10-080-505-13  
Sequence 13; Application US/10080505  
Publication No. US2003007316CA1

GENERAL INFORMATION:

APPLICANT: St. Gene, Joseph W.

TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS

FILE REFERENCE: A-59941-1/RT/DCF/DHR

CURRENT APPLICATION NUMBER: US/10/080,505

PRIOR FILING DATE: 2002-02-22

PRIOR APPLICATION NUMBER: US 08/296,791

PRIOR FILING DATE: 1994-10-25

PRIOR APPLICATION NUMBER: US 09/839,996

PRIOR FILING DATE: 2001-04-20

NUMBER OF SEQ ID NOS: 58

SOFTWARE: Patent in version 3.1

SEQ ID NO 13

LENGTH: 1436

TYPE: PRT

ORGANISM: Haemophilus influenzae

US-10-080-505-13

Query Match 6.3%; Score 112.5; DB 4; Length 1436;

Best Local Similarity 21.1%; Pred. No. 2.8; Mismatches 103; Indels 97; Gaps 14;

Matches 64; Conservative 39; Mismatches 103; Indels 97; Gaps 14;

4 REDSMDSMTGVSTFN-----VILETGDNINIVANGLMIGVINIT----- 45

570 KEIAYNGWFETDENKNGNLNIYKPTTDRITLLSGGTNLKG--NITQEGGTLVFSG 626

46 -PTDDEGNFVD-----IDVTINDNIKIYDY-IDG-----SDIDGSDG 81

627 RPTPHAYNHLNRNELGRPGQEVVIDDWTTRTFKAFNFOIKGSAVSVNSVSIEN-- 684

82 WFTYGNPN-EYNTIPNSQSYSLKSENSQITQIKRYVCSNTSLRTKSPSAKYTTTSGK 140

685 WTVSNNAANAFAVVPNOQNTICTRSDWTGLT-----TCKTVDLTDKVINSLIPTQING 738

141 VISITONS-----INSSRVIN-----AIDATN----- 163

739 SIMLTDNATYVINGLAKNGVTLINHSOPTLSNNATQTCGNISNNAATVDNANLNGN 798

164 --FTDDELRTTKETRFENQSYSHKSTNSLYVHTWTIPRSLKIQNRWEDYNNGWMTAQ 221

799 VHLTDSAQFSLKNSHFSHQ-IQGDKDTVTYVLENATWTMPSDATLQNL---TLNNSVTYTLN 854

222 SCY 224

855 SAY 857

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

Search completed: January 30, 2006, 09:52:52  
Job time : 22.8022 secs

RESULT 15  
US-10-687-046-13  
Sequence 13; Application US/10687046  
Publication No. US20040157241A1

GENERAL INFORMATION:

APPLICANT: St. Gene, Joseph W.

TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS

FILE REFERENCE: A-59941-1/RT/DCF/DHR

CURRENT APPLICATION NUMBER: US/10/687,046

PRIOR FILING DATE: 2003-10-15

PRIOR APPLICATION NUMBER: US/10/080,505

PRIOR FILING DATE: 2002-02-22

PRIOR APPLICATION NUMBER: US 08/296,791

PRIOR FILING DATE: 1994-10-25

PRIOR APPLICATION NUMBER: US 09/839,996

PRIOR FILING DATE: 2001-04-20

NUMBER OF SEQ ID NOS: 58

SOFTWARE: Patent in version 3.1

SEQ ID NO 13

LENGTH: 1436

TYPE: PRT

ORGANISM: Haemophilus influenzae

US-10-687-046-13

Query Match 6.3%; Score 112.5; DB 4; Length 1436;

Best Local Similarity 21.1%; Pred. No. 2.8; Mismatches 103; Indels 97; Gaps 14;

Matches 64; Conservative 39; Mismatches 103; Indels 97; Gaps 14;

4 REDSMDSMTGVSTFN-----VILETGDNINIVANGLMIGVINIT----- 45

570 KEIAYNGWFETDENKNGNLNIYKPTTDRITLLSGGTNLKG--NITQEGGTLVFSG 626

46 -PTDDEGNFVD-----IDVTINDNIKIYDY-IDG-----SDIDGSDG 81

627 RPTPHAYNHLNRNELGRPGQEVVIDDWTTRTFKAFNFOIKGSAVSVNSVSIEN-- 684

82 WFTYGNPN-EYNTIPNSQSYSLKSENSQITQIKRYVCSNTSLRTKSPSAKYTTTSGK 140

685 WTVSNNAANAFAVVPNOQNTICTRSDWTGLT-----TCKTVDLTDKVINSLIPTQING 738

141 VISITONS-----INSSRVIN-----AIDATN----- 163

739 SIMLTDNATYVINGLAKNGVTLINHSOPTLSNNATQTCGNISNNAATVDNANLNGN 798

164 --FTDDELRTTKETRFENQSYSHKSTNSLYVHTWTIPRSLKIQNRWEDYNNGWMTAQ 221

799 VHLTDSAQFSLKNSHFSHQ-IQGDKDTVTYVLENATWTMPSDATLQNL---TLNNSVTYTLN 854

222 SCY 224

855 SAY 857

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

THIS PAGE LEFT BLANK

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2006, 09:46:19 ; Search time 2.66268 Seconds  
(without alignments)  
1358.372 Million cell updates/sec

Title: US-09-889-874A-22

Perfect score: 1784  
Sequence: 1 FTLRDMSMTGVTSTFNV.....ILFHAGVNSDPYLSSTIV 334

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 75621 seqs, 10829074 residues

Total number of hits satisfying chosen parameters: 75621

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

Published Applications AA New:  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/PTC\_NEW\_PUB pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	118	6.6	773	US-11-052-554A-341	Sequence 341, App
2	107	6.0	1449	US-11-052-554A-237	Sequence 237, App
3	101	5.7	3194	US-11-052-554A-90	Sequence 90, App1
4	100.5	5.6	331	US-11-016-564-3	Sequence 3, App1
5	100.5	5.6	829	US-10-880-144-6	Sequence 6, App1
6	99.5	5.6	2399	US-11-052-554A-92	Sequence 92, App1
7	99.5	5.6	2902	US-11-052-554A-91	Sequence 91, App1
8	96.5	5.4	744	US-11-052-554A-94	Sequence 94, App1
9	95.5	5.4	807	US-11-052-554A-55	Sequence 55, App1
10	95.5	5.4	1468	US-10-467-657-1088	Sequence 1088, App
11	95	5.3	980	US-11-052-554A-17	Sequence 17, App1
12	94	5.3	2314	US-11-013-759-11	Sequence 11, App1
13	93.5	5.2	486	US-10-416-047-14	Sequence 14, App1
14	93.5	5.2	710	US-11-089-551A-33	Sequence 33, App1
15	92.5	5.2	1579	US-11-052-554A-9	Sequence 9, App1
16	92.5	5.2	2053	US-11-013-759-9	Sequence 9, App1
17	91.5	5.1	453	US-11-052-554A-224	Sequence 224, App
18	91.5	5.1	539	US-11-069-642-16	Sequence 16, App1
19	91.5	5.1	795	US-11-052-554A-87	Sequence 87, App1
20	91.5	5.1	826	US-10-793-626-1066	Sequence 1066, App
21	91	5.1	2340	US-11-052-554A-171	Sequence 171, App
22	90	5.0	395	US-10-793-626-668	Sequence 668, App
23	90	5.0	1250	US-11-052-554A-16	Sequence 16, App1
24	89	5.0	2715	US-11-113-424-51	Sequence 51, App1
25	88.5	5.0	566	US-11-033-039-1244	Sequence 1244, App

26	88.5	5.0	627	US-10-873-528-191	Sequence 191, App
27	88	4.9	1461	US-11-052-554A-283	Sequence 283, App
28	87	4.9	523	US-10-641-678-45	Sequence 45, App1
29	87	4.9	952	US-11-207-626A-45	Sequence 45, App1
30	86.5	4.8	518	US-11-052-554A-355	Sequence 355, App
31	86.5	4.8	1770	US-11-103-957-21	Sequence 21, App1
32	86	4.8	477	US-11-089-551A-34	Sequence 34, App1
33	86	4.8	930	US-10-517-939-106	Sequence 106, App
34	85.5	4.8	878	US-11-052-554A-23	Sequence 23, App1
35	85.5	4.8	1067	US-11/062	Sequence 3, App1
36	85.5	4.8	1092	US-11/062	Sequence 6, App1
37	85.5	4.8	1694	US-11-052-554A-83	Sequence 83, App1
38	85.5	4.8	3623	US-10-995-561-593	Sequence 593, App
39	85	4.8	1094	US-10-517-939-144	Sequence 144, App
40	85	4.8	2376	US-11-036-051-4	Sequence 4, App1
41	85	4.8	2715	US-11-036-051-2	Sequence 2, App1
42	84.5	4.7	338	US-11-016-564-11	Sequence 11, App1
43	84	4.7	295	US-10-467-657-5222	Sequence 5222, App
44	84	4.7	857	US-11-052-554A-218	Sequence 218, App
45	83.5	4.7	587	US-11-052-554A-118	Sequence 118, App

#### ALIGNMENTS

RESULT 1  
US-11-052-554A-341  
; Sequence 341, Application US/11052554A  
; Publication No. US2005028866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; CURRENT FILING DATE: 2005-02-07  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; PRIOR FILING DATE: 2004-02-06  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 341  
; LENGTH: 773  
; TYPE: PRT  
; ORGANISM: Shigella flexneri 2a str. 2457T  
; US-11-052-554A-341

Query: Match 6.6%; Score 118; DB 7; Length 773;  
Best Local Similarity 22.7%; Pred. No. 0.032;  
Matches 83; Conservative 55; Mismatches 151; Indels 76; Gaps 18;  
QY 2 TLNRDMSMTGVTSTFNVILT-----GLDNCNIYANGIMIGVITNPTDDEG 51  
DB 76 SLDBESTAD-TGSNNENNAIAKDMAGEITTHGESYAAN-----GTVAKGDTLDYT 129  
QY 52 N----FVDIDVTILNDINIKVYIDGSDIDSGMFTTGNPNR-----YN---TINSQ 98  
DB 130 NASVTLLDVDTITGDAHAIAAROGT-VFENGEBIYTTGPDAIAKIYGVGTVTLNNTS 188  
QY 99 SYS-----LKSNSQIQIKRYVSCNTSRLT-----KSPFAKTTSGKVISI 144  
DB 189 AVAHQSGIYLESSINQ---EATVDILSGSSIRKSNELIYHKDETISNVTITDSEVSA 244  
QY 145 TONSINSSRVINAIDATNFTDELRITKTERENOSYTSKSTNSLYVHTWTIPSLK 204  
DB 245 ADVFINNIKHL-TVATN---SKTGSANISITDNDHTYLSDNS---TWDIKDST 296  
QY 205 LQWRMEDVYNGMTMAOSCTYKTGADGGS-ESTRWLAAGSIFPPGNYDGLMDNDILSG 263  
DB 297 VSN-----LTVNSTVYISRADGRDVEPRLTITENYV---GNNGVILRLTELDNDN 345  
QY 264 MAKSYVVDVGINQLSPFTRILIGKGFVWVNIISGLDRGHAVIIDIQGN-----KRIIEFA 319

```
Db      346 SAIDKVINNTSGTTRVKTVMAGSSAYTLNGLIE-----ITSVGESENGEPIIDSRIFAG 401
Qy      320 GYENS 324
Db      402 AYEYS 406

RESULT 2
US-11-052-554A-237
; Sequence 237, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; PRIOR FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 237
; LENGTH: 1449
; TYPE: PRT
; ORGANISM: Neisseria meningitidis Z2491
US-11-052-554A-237

Query Match      6.0%; Score 107; DB 7; Length 1449;
Best Local Similarity 21.1%; Pred. No. 0.54; Mismatches 126; Indels 94; Gaps 17;
Matches 71; Conservative 46;

Qy      9 SDMTGVSTFVYLTGDLNCRIVANGLMIGVIINITPTDDEGNFVDID-----V 59
Db      722 SDWVGLT--NCEVETITDD-----KVIALSTKTDISGN-VSLADHAHLNLGLA 767
Qy      60 TLNDNIKIVDYIDSDIDSDGSMFYTGPNPEYNTIPN-SQSYSLKSENSQITQIKYVS 118
Db      768 TLNGLLS-----ANGDRTYVSHANTONGNLSLVGNAQAT-FNQATL 808
Qy      119 CSNSTRLRKTSFSAKVTITTSKVISITONS--INSRVVINAIDATNFTDDELRTTKETRF 177
Db      809 NGNTSASGNAASFJNSNNAONGSLTLDNAKANVSHSLNCG-----NVSLLAKAVFHF 861
Qy      178 ENOSYTSKHSSTNSLYVH-----TWTPRSKLQVWREDDYNNGWMTWASQCY--KTGADG 231
Db      862 ENSRFTQGLSGSKDTAHLKXOSEWTLPGCTELGNL--NLDNATITLNSAVRHDAAGQT 918
Qy      232 GS-----ESTRWLAAGSITPPGNYDGLMDLNDIALSGMAHKSYNVDGTGINSFTRII 284
Db      919 GSVSDTPRRRRSRRL--SVTPPTSVERF--NTLTVNGK-----LNGGCTFRFM 964
Qy      285 GKGFSSW-----VYNISGLDGHAVIITDQ 308
Db      965 SELFGYRSDKLAESSEGYTTLAVNNTGNRPVSIHQ 1001

RESULT 3
US-11-052-554A-90
; Sequence 90, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; PRIOR FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
```

```
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 90
; LENGTH: 3194
; TYPE: PRT
; ORGANISM: Helicobacter pylori J99
US-11-052-554A-90

Query Match      5.7%; Score 101; DB 7; Length 3194;
Best Local Similarity 18.6%; Pred. No. 4.4;
Matches 84; Conservative 51; Mismatches 140; Indels 176; Gaps 22;

Qy      12 TGUSTVNVILETGLDNCIYA--NG-----LNMIGVI--INTPTDDEGNFVIDDVTL 61
Db      561 TGFTFADRVYITG---NMWISGNGAQTGGGATNPGVATEINIA---GATPKALKTTTQ 612
Qy      62 ND-----NIKIVDYIDSDIDSDGSMFYTG----- 87
Db      613 NSYMTFVALONGSGSGKINVSQSDFYDWD---GGYDFGNGVFDPSVNFNKAYYKFGA 668
Qy      88 PNEYN-----TIPNSQSYSLKSENSQITQIK-----RIVSCS 120
Db      669 ENSYNEFKNTFLAGNFQCKTTEKSVLNDASYA--FDGVNNAFNEDEKFGSGSEFNHAB 727
Qy      121 NTSRLRTRKTSFSAKVTITTSKVISITONSINSRVVINAIDATNFTDDELRTTKETRF-- 178
Db      728 QTNAPNNNSPFGSGSFSAKQVDPNGNSFNGVAFNFTNTKASFTINDTFVNNQFKLNGA 787
Qy      179 -----NOSYTSKHSSTNSLY-VHTW----- 197
Db      788 QTDFTFSKGVFNMQGLLSLSVGTTYQLNAKSQVKONNMNLQMLRWTSGENPSGKL 847
Qy      198 -----TTPRSKLQVWREDDYNNGWMTWASQCYTKTGADGSESTRWLAAGSITPPGNYDG 252
Db      848 VDEKTAIPNSAKIYNOQFTD--NGLTY---YIKENFNNGIITLRLCTLGYTHCVN--- 897
Qy      253 LMLDNDIALSGMAHKSYNVDGTGINSFTRIIIGKFSFVYNNISGL-----DRGHAVIITDQ 308
Db      898 --IDND-----ANLKNVNNNANSNT-----VYLNQMTTWKTRAGTGVFTQDY 937
Qy      309 YGNKYRIILFHAGYENSDDPYL-----SSSIV 333
Db      938 SGTNSVAVFN---QTTPLFAGNPTNSNVV 964

RESULT 4
US-11-016-564-3
; Sequence 3, Application US/11016564
; Publication No. US20050287146A1
; GENERAL INFORMATION:
; APPLICANT: INHIBITEX, INC.
; TITLE OF INVENTION: METHOD OF INHIBITING CANDIDA-RELATED INFECTIONS
; FILE REFERENCE: P08140US03/BAS
; CURRENT APPLICATION NUMBER: US/11/016,564
; PRIOR FILING DATE: 2004-12-20
; PRIOR APPLICATION NUMBER: US 60/566,082
; PRIOR FILING DATE: 2004-04-29
; PRIOR APPLICATION NUMBER: US 60/561,540
; PRIOR FILING DATE: 2004-04-13
; PRIOR APPLICATION NUMBER: US 60/530,654
; PRIOR FILING DATE: 2003-12-19
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Candida albicans
US-11-016-564-3

Query Match      5.6%; Score 100.5; DB 7; Length 331;
Best Local Similarity 25.5%; Pred. No. 0.3;
```

Matches 42; Conservative 23; Mismatches 55; Indels 45; Gaps 8;  
QY 9 SDWGTSTFVILETGLDNCNIYANGLMG--VIINITPTDDEGNFVDIDDT-----60  
Db 165 SMFTNIDTKHATVEQ-----TIYINPLRSYAKETNIVISGEGESTI-IDSTIIRKYK 218  
QY 61 -----LNDNIKIVDYIDGSDIDGDMFYTGPNNEVNT-----IPNS 97  
Db 219 VGDNGNLPDSNRITDYSEYEDVT-NDIYAQLGNNDVNIINFGINDSPYIIKVISKIDPNK 277  
QY 98 QSYSLKSENSQITQIKRYVCSNTSRLRTKFSKAV--TTTSGK 140  
Db 278 DDTYTIQCTVTMGTITNEY-----TGEFTASYDNTIAFTSSGQ 317

RESULT 5  
US-10-880-144-6  
; Sequence 6, Application US/10880144  
; Publication No. US20050287537A1  
; GENERAL INFORMATION:  
; APPLICANT: Svendsen, Allan  
; TITLE OF INVENTION: Pullulanase variants and methods for preparing such variants with  
; FILE REFERENCE: 6072.220-US  
; CURRENT APPLICATION NUMBER: US/10/880,144  
; NUMBER OF SEQ ID NOS: 40  
; SOFTWARE: Patentin version 3.3  
; SEQ ID NO 6  
; LENGTH: 829  
; TYPE: PR1  
; ORGANISM: Bacillus acidophilus  
US-10-880-144-6

Query Match 5.6%; Score 100.5; DB 6; Length 829;  
Best Local Similarity 21.5%; Pred. No. 0.92;  
Matches 71; Conservative 44; Mismatches 106; Indels 109; Gaps 15;  
QY 11 WTGVSFNVILETGLDNCNIYANGLMG--IGVINITPTDDEGNFVDIDDTLNDNIKI 67  
Db 551 WTGCT-----GGLSSDQLVTKGQKGLIGV-----FNNIR- 582  
QY 68 VDYIDGSDIDGDMFYTGPNNEVNTIPN-----SOSYSLKSENSQITQIK 114  
Db 583 -NGLDGNVPRKSAQGFATGDPNOVNIKRVMSISDFTSPABETIIVYSHDNMTLMDK 641  
QY 115 RYVSCSN-----TSRLTKSFSKAVTTTS-----GKVISITONSINSSRVVINAIDA 161  
Db 642 --ISASNPNDTQADRIMDELAQAVVETSGVPMGGSEMLRTKGNDS--YVAGDS 696  
QY 162 TNETDELRTTKTRPENOSYTS-----KSTNSLYVHTWTIPRSLK- 204  
Db 697 VNOFDSRKQAFENVVDYVSMWLHLRDNHAPFMTTADQIKOMLTFDSPTNVAARELKN 756  
QY 205 -LQNMWED-----YNNGMWAGSCYKTAGDGSESTRMLAASIRPGVYDGLMDNDI 259  
Db 757 HANHDKKKILIVVYNNPKT-----AQLTTLPSGMMTIVGLGNV 795  
QY 260 ALSGMAKSYNVDGINQSLFTRIIGKFS 289  
Db 796 GEKSLGHVNGTVE--VPALS-TIILHGTS 822

RESULT 6  
US-11-052-554A-92  
; Sequence 92, Application US/11052554A  
; Publication No. US20050288866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A

; CURRENT FILING DATE: 2005-02-07  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; PRIOR FILING DATE: 2004-02-06  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: Patentin version 3.3  
; SEQ ID NO 92  
; LENGTH: 2399  
; TYPE: PR1  
; ORGANISM: Helicobacter pylori J99  
US-11-052-554A-92

Query Match 5.6%; Score 99.5; DB 7; Length 2399;  
Best Local Similarity 23.0%; Pred. No. 4.1;  
Matches 75; Conservative 38; Mismatches 128; Indels 85; Gaps 12;  
QY 13 GVSTFNVILETGLDNCNIYANGLMIGVINITPTDDEGNFVDID-----DVLNDNIK 66  
Db 655 GVANNGTLLIGNTSVNDNGLIWI-----HGFYITGTGFSANITYLTNFK 704  
QY 67 IVDYIDGSDIDG-----DGWFY-----TG-NPNEVNTIPNSQSYSL 102  
Db 705 TGEVGSNDGCGANITFKASDNITMDQLNNAETVKMIQTGASGHSYTFPATNNISV 764  
QY 103 LKSENSQITQIKRYVCSNTSRLRTKFSKAVTTTSGKVISITONSINSSRVVINAIDA 162  
Db 765 TDSDFSDMTWGK-----PSFSAKISFSNAPSGFTNCGSSTISTNANSL 811  
QY 163 NPTDDELK-----TTKETPENOSY-----TSKSTNSLYVHTWTIPRSLKQ 206  
Db 812 SFTDSRLNGALYNLOANSLIFNNTOAVRVVLYSRGTSNPNATLTGLNTSPLSSQSL 871  
QY 207 NMRMEDYNNGMWAGSCYKTAGDGSESTRMLAASIRPGVYDGLMDNDIALSGMAR 266  
Db 872 NF-----NGDTLQK--NANITLGNKSG--AAFNSLTLDNVNSLSDNOSVLANGT 920  
QY 267 KSYNVDTGIN-----OLSTRIIGK 287  
Db 921 SAFNQAASINITYNGSAARFSLFNG 946

RESULT 7  
US-11-052-554A-91  
; Sequence 91, Application US/11052554A  
; Publication No. US20050288866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; CURRENT FILING DATE: 2005-02-07  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; PRIOR FILING DATE: 2004-02-06  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: Patentin version 3.3  
; SEQ ID NO 91  
; LENGTH: 2902  
; TYPE: PR1  
; ORGANISM: Helicobacter pylori J99  
US-11-052-554A-91

Query Match 5.6%; Score 99.5; DB 7; Length 2902;  
Best Local Similarity 17.9%; Pred. No. 5.1;  
Matches 66; Conservative 63; Mismatches 146; Indels 93; Gaps 14;  
QY 15 STFNVILETGLDNCNIYANG--LNNIGVINITPTDDEGNFV-----DI 56  
Db 145 ATFNLAAGSNGSFTSWYPNGHTDVTSPAGTIIVNNSVEVGNRVGSGAGTGTATLNLNA 204

[illegible]

```

RESULT 8
US-11-052-554A-94
; Sequence 94, Application US/11052554A
; Publication No. US2005028866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: Patentin version 3.3
; SEQ ID NO 94
; LENGTH: 744
; TYPE: prt
; ORGANISM: Helicobacter pylori J99
US-11-052-554A-94

```

[illegible]

RESULT 9  
US-11-052-554A-55  
; Sequence 55, Application US/11052554A  
; Publication No. US2005028866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL

```

? FILE REFERENCE: 30853/40359A
? CURRENT APPLICATION NUMBER: US/11/052,554A
? CURRENT FILING DATE: 2005-02-07
? PRIOR APPLICATION NUMBER: US 60/589,227
? PRIOR FILING DATE: 2004-07-20
? PRIOR APPLICATION NUMBER: IN 173/DEL/2004
? PRIOR FILING DATE: 2004-02-06
? NUMBER OF SEQ ID NOS: 763
? SOFTWARE: PatentIn version 3.3
? SEQ ID NO 55
? LENGTH: 807
? TYPE: PRT
? ORGANISM: Escherichia coli 0157:H7
? US-11-052-554A-55

```

Query Match	5.4%	Score 95.5;	DB 7;	Length 807;
Best Local Similarity	20.4%;	Pred. No.2.3;		
Matches	68;	Conservative	38;	Indels 101; Gaps 15.
Dd	15	STFVNIETGLDNCNIYANG-----INMLGVIINTPTDEGNFVIDIDVTINDNIKIV	68	
	:	: : :	:	:
Db	269	ATAVAIEIEN--GTINIFRANDEFARSVLTGBEHIAN-----NGRIVALDGVTGGSLGIK	321	
	:	: : :	:	:
Qy	69	DYIDSGSDIGSDG-----WPTYGNPNEYNTIPNSQSYSLLKSENSOITQIKRYISC	119	
	:	: : :	:	:
Dd	322	DSVANVEGVNGSNNTEHVHTDYTLPLMPFNPTYTTSPESETTDSGSDDSS--NNINGXYIVG	380	
	:	: : :	:	:
Qy	120	SNTSRRLTKRSASAVYTTSGKVIIITQNSINSRRVVINAIDATNPFDDELTKKETREPN	179	
	:	: : :	:	:
Dd	381	TN-----VDGSAGK--LKVNNAAMNGGIN---TGPAAGTADITT--VSFDN	419	
	:	: : :	:	:
Qy	180	QSYTHKSSTNSLYVHTWTPRPRLKLQWRMEDYNNCGWTMAOSCYUKTAGDG--GSESTRW	238	
	:	: : :	:	:
Dd	420	-----VVEGINL-----TDDAISTSVVM	439	
	:	: : :	:	:
Qy	239	LAAGSIFFPGNYDGLMNDIDIALSGMAHKSYNDVTGINQLSFTRIIGKGFSWYNISGLD	298	
	:	: : :	:	:
Dd	440	TAKGSTDASGNV-----DVIMSKNAATVDATDASVNDVA-----KALD	477	
	:	: : :	:	:
Qy	299	RGHAVIITIQQYGNKRRIILFHAGYENSDDPYLSSI	332	
	:	: : :	:	:
Dd	478	-----AGYTNNELYTSLNV	491	
	:	: : :	:	:

```

RESULT 10.10-657-1088
; US-10-467-1088
; Sequence 1088, Application US/10467657
; Publication No. US2005260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqMan99, version 1.04
; SEQ ID NO 1088
; LENGTH: 1468
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-1088

```

```

Query Match          5.4%; Score 95.5; DB 6; Length 1468;
Best Local Similarity 22.6%; Pred. No. 4.7;
Matches 63; Conservative 36; Mismatches 109; Indels 71; Gaps 15

```

Db 729 SDWTGLSCRE-KTIYDD-----KVIAISLKTIDRGN-VSLADHAHLNLTGIA 774  
Qy 60 TLNDNIKIYIDISDIDSGDMFYTNPNREYNTIP-SQSYSLKENSQITQIKRYVS 118  
Db 775 TLNGLNIS-----ACGDHYTYTRATQNGNLSLVGNAQAT-FNQATL 815  
Qy 119 CSNTRSRKTSFSAKVTTSKGVISITONS-INSSRVINAIADINTFTDELRTKTRF 177  
Db 816 NGNTSASDNASFNLSNNAVONGSLTLDNAQANVSHALNG-----NVSADKAVFHF 868  
Qy 178 ENOSYTSKSTNSLYVH-----TWTPRSLKQWNRMEDYNNGTWMAQSCYY-KTGADG 231  
Db 869 ENSFTKISGKOTATLHLDSEWTLPSGTELGTL---NIDNATITLNSAYRHDAQAQOT 925  
Qy 232 GS-----ESTRWLAAGSIFPPGNYDGLWLDNDIALSG 263  
Db 926 GSADAPRRRSRSL--SVTPPTSASRF--NTLTVNG 960

## RESULT 11

US-11-052-554A-17  
; Sequence 17, Application US/11052554A  
; Publication No. US20050288866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; CURRENT FILING DATE: 2005-02-07  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; PRIOR FILING DATE: 2004-02-06  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 17  
; LENGTH: 980  
; TYPE: PRT  
; ORGANISM: Escherichia coli 0157:H7  
US-11-052-554A-17

Query Match 5.3%; Score 95; DB 7; Length 980;

Best Local Similarity 19.8%; Pred. No. 3.2;  
Matches 68; Conservative 55; Mismatches 108; Indels 112; Gaps 18;

Qy 5 EDMSMDWTGVSTFVILETGLDNCNY-----ANGLMIGVITINTPTD 48  
Db 147 DDNYEHYTYNGITNNADTHV--VDVYDMGTAITLDEVDLSITNNSHVAGITLT----- 199  
Qy 49 DEG-NFVDIDVTL-----NDNIKIYD-YIDGSDID-GSDGMF-YTGNPNREYNT 93  
Db 200 -QGWEYEDIDNTVSTSVNSSEVENNTITVDSFTVSGWTDGEGTGMFGHTGASNY-- 256  
Qy 94 IRPSQSISLKSNSQITQIKRYVSCNTRSRKTSFSAKVTTSKGVISITONSINSR 153  
Db 257 -----SNTLTADVAIAAANPYADNMQTTY-TLDNST 289  
Qy 154 VINAIADATNFTDELRTKTRPENOSYTSKSTNSLYVHTWTPRSLKQWNRMEDY 213  
Db 290 LMGDVVPSSNFDER-----FFPGANSYRPAADDVUTNGMDGDRBDV-----TL 334  
Qy 214 NNG--WTMAQSCYYKTGADG-----GSESTRWL--AAGSIFPPG-----NY 250  
Db 335 NNGSKWGAAMSYMWDEDDGSYDGYAVGTGATATLIDIAANSLWPSSTVGVNDINTQY 354  
Qy 251 D--GLWLDNDIALSGMAHKSYN-----VDTGINDLSFRIITGK 287  
Db 395 DENGHIVNEVYQGLFNVTILNGSEWDYTKSSLDITLINSNG 437

## RESULT 12

US-11-013-759-11  
; Sequence 11, Application US/11013759  
; Publication No. US20050249747A1  
; GENERAL INFORMATION:  
; APPLICANT: Loomore, Sheena M.  
; APPLICANT: Sasaki, Ken  
; APPLICANT: Yang, Yan Ping  
; APPLICANT: Klein, Michel H.  
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE  
; FILE REFERENCE: 1038-921MIS:jb  
; CURRENT APPLICATION NUMBER: US/11/013,759  
; CURRENT FILING DATE: 2004-12-16  
; PRIOR APPLICATION NUMBER: US/09/361,619  
; PRIOR FILING DATE: 1999-07-27  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 11  
; LENGTH: 2314  
; TYPE: PRT  
; ORGANISM: Moraxella catarrhalis  
US-11-013-759-11

Query Match 5.3%; Score 94; DB 7; Length 2314;

Best Local Similarity 22.0%; Pred. No. 11;  
Matches 68; Conservative 44; Mismatches 103; Indels 94; Gaps 17;

Qy 44 ITPTDEGNFVDIDV-----TLNDNIKIYIDISDIDSGDMFYTNPNREYNTIP--- 95  
Db 748 ITSDDEKSKAASIGDILNTGFNLKNNNSVGFV-----STNTVTFID 790  
Qy 96 -NQSYSLKENSQITQIKRYVSC-----SNTSRKTSFSAKVTTSKGVIS 143  
Db 791 GNAITAKVYDEFNQTSKYTYDVAVDEKTELTGDNKTKIKGVKTTTLTTNANGKA-- 848  
Qy 144 ITONSINSRVINAIADAT--NFTDELRTKTR-----RFENOSYTSKH 186  
Db 849 -TNFSTINDNALVNAQDIANLNTLAKEHITTTGTADTLQTRKVKDQATDETTIVGK 907  
Qy 187 SST-NSLYHTWTPRSLKQWNRMEDYNNGTWMAQSCYYKTGADG--GSESTRWLA 241  
Db 908 DGTONGKTVNT-----LTKKG-----ENGLTVA-----TKKDGTVTFGINTQSLKA 949  
Qy 242 GSIFPPGNYDGLWLDNDIALSGMAHKSYNVDTGINDLSFRIITGKGSWYINISGLDRGH 301  
Db 950 GD-STLNDGSLSKIPAS-----NEQTVGADGVKFAK-VDKNS--STGID-GT 995  
Qy 302 AVTIIDQYV 310  
Db 996 SRITKQDIG 1004

## RESULT 13

US-10-416-047-14  
; Sequence 14, Application US/10416047  
; Publication No. US20050266512A1  
; GENERAL INFORMATION:  
; APPLICANT: Buckley  
; TITLE OF INVENTION: Detection of Proteases and Screening for Protease Inhibitors  
; FILE REFERENCE: 63198-1353  
; CURRENT APPLICATION NUMBER: US/10/416,047  
; CURRENT FILING DATE: 2003-05-05  
; PRIOR APPLICATION NUMBER: 60/247,160  
; PRIOR FILING DATE: 2000-11-07  
; PRIOR APPLICATION NUMBER: PCT/CA01/01561  
; PRIOR FILING DATE: 2001-11-07  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 14  
; LENGTH: 486  
; TYPE: PRT  
; ORGANISM: Aeromonas hydrophila  
US-10-416-047-14

Query Match 5.2%; Score 93.5; DB 6; Length 486;  
Best Local Similarity 19.1%; Pred. No. 1.8;  
Matches 66; Conservative 41; Mismatches 108; Indels 131; Gaps 15;  
QY 8 MSDW--TGSTFNVILETGLDN-----CNIVANGLMI--GVINITPTD--DEGNFVDI 56  
DB 65 MGWQISGLANGVIMPGNGELKPGTASNTWCYFPNPVTGELPILSALDIDGDEVDV 124  
QY 57 DDTVLNDN--IKIVDYID-----GSDID--GSDGPFYTGPN-----IKRY 116  
DB 125 QMRLVHDSANFIKPTSYLAHYLGAVWGNHSGYVGEDMDVTDGDMVIRGNNDGCGDG 184  
QY 90 -----EYVTINSOSYSLKSENGIITG-----IKRY 116  
DB 185 YRCGDKTAIRVSNFAYNLDPDSFPGHGVDTQSDRQLVKTGVVAVNDSDPTQSGYDTLRY 244  
QY 117 VSCSNTSRLTKTSPSAKYTTTS--GKVISITONS--NSRVINAIIDATNFTDEL 169  
DB 245 DTATNMSKNTYGLSEKVTTKNKKFMPVNGETQSLIANGSASQNGSTTSLSGSV 304  
QY 170 RTTKETR-----FENQSYTSH-----KSSTNSLYVH-----TWT 198  
DB 305 RPTVPASAKIPVKIELYKADISYEFKADVSYDLTSGFLRMGNAMYTTHPDNRPMWNH 364  
QY 199 -----IPRSLKQNRWEDYNNGWMAOS 222  
DB 365 TFIYGPYKAKASIRYQMDKRYIPGEVAKWMDMNTTIOQNGLSTWON 410

RESULT 14  
US-11-089-551A-33  
; Sequence 33, Application US/11089551A  
; Publication No. US20050266242A1  
; GENERAL INFORMATION:  
; APPLICANT: Lindquist et al.  
; TITLE OF INVENTION: ELECTRICAL CONDUCTORS AND DEVICES FROM PRIOR-LIKE PROTEINS  
; FILE REFERENCE: 30554/40025A  
; CURRENT APPLICATION NUMBER: US/11/089,551A  
; PRIOR FILING DATE: 2005-03-24  
; PRIOR APPLICATION NUMBER: US 60/559,286  
; NUMBER OF SEQ ID NOS: 65  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 33  
; LENGTH: 710  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-11-089-551A-33

Query Match 5.2%; Score 93.5; DB 7; Length 710;  
Best Local Similarity 22.8%; Pred. No. 2.8;  
Matches 41; Conservative 31; Mismatches 85; Indels 23; Gaps 5;  
QY 70 YIDSDIDGSDGWFYTGPNREYNT-----IPNSQSYSLKSENSQTOIKRYVSCNNTS 123  
DB 6 YNGNSNVDTG-----YDRDRTMTNABSVRVSDEKGRSSSTKSGYRTRAGRSDTJNSA 61  
QY 124 RLRTKSPSAKYTTTSGKVISITONSINSSRVINAIIDATNFTDELRTTKETREPNOSYT 183  
DB 62 KHSKSKTIVVVVTS-----STDSNSTYARVSSDSTVATS-----STTTRTRTNNTVS 110  
QY 184 SHKSTNSLYVHTWITIPRSLKQNRWEDYNNGWMAOSCYKYGADGSGESTRWLAAGS 243  
DB 111 STASSSTTDVGNATSNANWSANASNTSSSDYATSTYRKSJTDNTYA--NSKNGNNSSAGN 168

RESULT 15  
US-11-052-554A-9  
; Sequence 9, Application US/11052554A  
; Publication No. US2005028866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.

; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; CURRENT FILING DATE: 2005-02-07  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; PRIOR FILING DATE: 2004-02-06  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 9  
; LENGTH: 1579  
; TYPE: PRT  
; ORGANISM: Escherichia coli 0157:H7  
US-11-052-554A-9

Query Match 5.2%; Score 92.5; DB 7; Length 1579;  
Best Local Similarity 18.9%; Pred. No. 9;  
Matches 68; Conservative 67; Mismatches 121; Indels 103; Gaps 18;  
QY 3 LRDSMSDWT--GVSTFNVILETGLDNCNI-----YANGLMI--GVINITPT--TDDEGN 52  
DB 415 VQDITLSDMTDNGDSGYTQILTAGTSSVTLTPQINGSAVRESIVNIVPVVSSRDHS 474  
QY 53 FVDIDDT--LNDNIKI--VDYIDSDIDGSDGWFYTGPNREYN-----TIPNSQS-- 99  
DB 475 SITIDNVSYAGDIAKRYELKDDSN-----QPAVYKQBELVKAVTENSXPGA 523  
QY 100 -----YSLKSENSQTOIKRYVSCN--TSRLTKTSPSAKYTTTSGKY--ISIT 145  
DB 524 TIWHEEQPGVYANYPAYKQGTALRAQLSIHWNNAPLQSHIYNEANQKARVATLSAT 583  
QY 146 QNSINSRVYIN--AIDATNFTDELRTTKETREPNOSYSH-----KSSTNSLYVHTWT 198  
DB 584 NNDVYADKTKFTNLTINVTDESNDPL--TNHQTFFKQKSGSAEFVPPQONTDAICVATIN 642  
QY 199 IPRSLKQNRWEDYNNGWMAOSCYKYGADGSGESTRWLAAG-----LSGMAKSYNVDTGI 242  
DB 643 MVSQVAEENITISATLPRNG--FSQRIIAKFPVSDSTPRFKQLVADPDTIANGSGSTLTA 700  
QY 243 -----SIFPGNYDGLMDNDIA-----LSGMAKSYNVDTGI 275  
DB 701 IITDFHNNPLKDMKVNVPAG--GSQLDNTTATTDGSGIVRVHLLTSKAGSYSVDAASL 756

Search completed: January 30, 2006, 09:53:18  
Job time : 3.66268 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: January 30, 2006, 09:41:43 ; Search time 5.82461 Seconds

(without alignments)  
5517.344 Million cell updates/sec

Title: US-09-889-874a-22

Perfect score: 1784  
Sequence: 1 FTLRDMSMDWTGVSTFENVL.....ILFHAGYENSDFYSSSIYV 334Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 200000000Post-processing: Minimum Match 0%  
Listing first 45 summaries

## Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	124	7.0	1092	2 T30214	fibrinogen-binding
2	122.5	6.9	990	2 E90433	conserved hypothe
3	117.5	6.6	640	2 B32935	hypothetical prote
4	115.5	6.5	1671	2 S71628	sensory transducti
5	111	6.2	369	2 F95107	glutamate 5-kinase
6	111	6.2	376	2 H97975	glutamate 5-kinase
7	109	6.1	807	2 B71605	hypothetical prote
8	109	6.1	1449	2 T30857	glucosyltransferas
9	109	6.1	1449	2 T30552	glucosyltransferas
10	108	6.1	626	2 T01485	probable polygalac
11	107.5	6.0	343	2 E70149	hypothetical prote
12	107.5	6.0	1038	2 H90053	hypothetical prote
13	107	6.0	1072	2 A86827	hypothetical prote
14	107	6.0	1114	2 JH0284	hypothetical prote
15	107	6.0	1449	2 B81963	125k surface serin
16	106	5.9	639	2 A32935	protein p1 - Entam
17	105.5	5.9	439	2 G97336	probable S-layer p
18	105	5.9	507	2 A71622	hypothetical prote
19	105	5.9	635	2 S57714	capB protein - Clo
20	105	5.9	667	2 E69733	PSX prophage ORF
21	105	5.9	1431	2 A45866	extracellular enzy
22	104	5.8	836	2 D97182	extracellular neut
23	104	5.8	1306	2 S25370	MSB2 protein - Yea
24	103.5	5.8	331	2 A69088	conserved hypothe
25	103.5	5.8	934	2 T39981	hypothetical prote
26	103.5	5.8	1031	2 A29839	RAD2 protein - Yea
27	103.5	5.8	1881	2 H95076	zinc metalloprotei
28	103	5.8	497	2 T47715	hypothetical prote
29	102.5	5.7	531	2 B96966	beta-mannanase Man

30	102.5	5.7	1148	2 S72635	exo-poly-alpha-gal
31	102	5.7	934	2 B29838	paraportal crystal
32	102	5.7	1224	2 E71611	hypothetical prote
33	102	5.7	2894	2 C64474	hypothetical prote
34	101.5	5.7	557	2 JC5487	cellulase (EC 3.2.
35	101	5.7	381	2 T28378	ORF MSV217 SCG gen
36	101	5.7	870	2 S27514	mosquitocidal toxi
37	101	5.7	903	2 T20804	hypothetical prote
38	101	5.7	1136	1 USB881	paraportal crystal
39	101	5.7	1457	2 D81019	adhesion and penet
40	101	5.7	3194	2 D71917	toxigen-like outer m
41	101	5.7	3972	2 S75251	hypothetical prote
42	100.5	5.6	1649	2 C86822	hypothetical prote
43	100	5.6	1577	2 A35140	hemolysin A precur
44	100	5.6	2329	2 T28125	hypothetical prote
45	99.5	5.6	962	2 AG2444	hypothetical prote

## ALIGNMENTS

```
RESULT 1
T30214
fibrinogen-binding protein - Staphylococcus epidermidis
C:Species: Staphylococcus epidermidis
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C/Accession: T30214
R:Nilsson, M.; Frykberg, L.; Flock, J.I.; Pei, L.; Lindberg, M.; Guss, B.
Infect. Immun. 66, 2666-2673, 1998
A>Title: A fibrinogen-binding protein of Staphylococcus epidermidis.
A:Reference number: Z20781; PMID:98261511; PMID:9596732
A:Accession: T30214
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1092 <NT>
A/Cross-references: UNIPROT:O70022; UNIPARC:UPI00000AFCa6; EMBL:Y17116; NID:e1296734; PIR
Query Match 7.0%; Score 124; DB 2; Length 1092;
Best Local Similarity 21.0%; Pred. No. 0.58;
Matches 78; Conservative 50; Mismatches 142; Indels 102; Gaps 16;
QY 9 SDWTGVSTFNVILETGLDNCNIYANGIMIG--VIINIFPTDDEGNFVDIDVT----- 60
DB 439 SMTNIDITKHYEQ-----TIINPLRYSAKTNVINSNGDEGSI--IDDSTIIRYVK 492
QY 61 -----LNDNIKIVDYIDSGSDGWFYTGPNEXNT-----IPNS 97
DB 493 VGDNQMLPDSNRIVDYSEYEDVT--NDVYQLGNNNDVINFGNIDSPYIIKVISKDPNK 551
QY 98 QSYSLKSENSQTOIKRYVSCNTSRIRKFSYSAKY--TTTSG-----KVISI-- 144
DB 552 DDYTTIQVTVMTTINEX-----TGEFRASYDNTIAFTSSGQGGDLPPERTYKIGD 606
QY 145 -----TQNSINSRVVINAIDATNFTDDELRTTKTRFENOSYTSKSKSTNSL-Y 193
DB 607 YWEDVDKQGIQNTNDEKLSNVLTLYTPDG---TSKSVRTDEBQKVPDGLKNGLT 663
QY 194 VHTWTIPRLSKLQNMWMBDY-----NNGMTWAGSCYYTGADGSEST 236
DB 664 KTFETP-----EGYTPILKISGTNPALDSRGNVWVTINGDDMTIDSGFYQTP 713
QY 237 RMLAAGSIFPPGVDGLMDNDIALSGMAKSNVDTGINSQSLFTNIIIGKFSWVNIISG 296
DB 714 KYSLGNYWYDNTKKDQIGDDEKIGSV--KVTLKDENGNIISTTTDENG---KYQPDN 768
QY 297 LDRGHAVIYIDQ 308
DB 769 LNSGNYIVHFDK 780
RESULT 2
E90433
conserved hypothetical protein [imported] - Sulfolobus solfataricus
```

C:Species: Sulfolobus solfataricus  
C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 09-Jul-2004  
C:Accession: E80433  
R:Site, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-  
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, H.  
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.  
submitted to GenBank, April 2001  
A:Description: Sulfolobus solfataricus complete genome.  
A:Reference number: A99139  
A:Accession: E80433  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-990 <KIR>  
A:Cross-references: UNIPROT:Q97VM0; UNIPARC:UPI0000647BD; GB:AEO06641; NID:G13815906; E  
C:Genetics:  
A:Gene: SSO2602

Query Match      6.9%; Score 122.5; DB 2; Length 990;  
Best Local Similarity 20.8%; Pred. No. 0.66;  
Matches 64; Conservative 45; Mismatches 99; Indels 99; Gaps 13;

OY 36 NMIGVINITPTDDEGNFVIDD-----VTINDNIIKYIDGSDIDSGDW 82  
| : |  
DB 728 NLVINISLNGKFTSN--NLIVNGPVSCLKHAYVLQYINLSNIPIKALINSTLVTFSPGW 785  
| : |  
OY 83 FYTGNP-----NEYTIIPNSQSYSLL-----KESNQITQIKRYVSCSNTSRRLRTX 128  
| : |  
DB 786 YNAVTLPISFIIMWTYYISNNTRTYIILSLPENFVTVNRSLTVAVKTLTKEL----- 834  
| : |  
OY 129 SPSAKVTTTSGKVSIITONSINSRRVYNALDAINFDDDLRTKEFRFKQSTSHKSS 188  
| : |  
DB 835 -----VTYNEBILKTIANRLTSEIWPA-----GQTLLPRKYNVISNNERIFYNTS 882  
| : |  
OY 189 TNSLYHTWTPRPRLKLQNMFMEDYNNGTWAQSCYYTGTAGDGSESTRWLAAQS----- 243  
| : |  
DB 883 S-----YLNIITOPTSINVK-----PIIEHYLTIDGSE---WLRPGSVTLTT 922  
| : |  
OY 244 ----TFPPGNVDGLMDNDIALSGMAHSYVDTGTINQLSFTRIIGFSPWVYNISGLDR 299  
| : |  
DB 923 QSVPIEGCGKWEK-----SYNVNGV-AITVNQPITETF--VKNIINGSFV 964  
| : |  
OY 300 GHAVIII 306  
| : |  
DB 965 GSVAIILI 971

RESULT 3  
B32935  
hypothetical protein NP1 - Entamoeba histolytica (fragment)  
C:Species: Entamoeba histolytica  
C:Date: 20-Dec-1989 #sequence\_revision 20-Dec-1989 #text\_change 09-Jul-2004  
C:Accession: B32935  
R:Taninich, E.; Horstmann, R.D.; Knobloch, J.; Arnold, H.H.  
Proc. Natl. Acad. Sci. U.S.A. 86, 5118-5122, 1989  
A:Title: Genomic DNA differences between pathogenic and nonpathogenic Entamoeba histolyt  
A:Reference number: A32935; MUID:8929655; PMID:2544890  
A:Accession: B32935  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-640 <TAN>  
A:Cross-references: UNIPROT:P20301; UNIPARC:UPI0000125CB8

Query Match      6.6%; Score 117.5; DB 2; Length 640;  
Best Local Similarity 21.9%; Pred. No. 0.83;  
Matches 73; Conservative 45; Mismatches 121; Indels 95; Gaps 18;

OY 25 LDNCNIYANG---LNMGVINITPTDDEGNFVIDDVTINDNIIKYIDYIDGS-----DI 76  
| : |  
DB 17 LDIRMSNDMRPIDSIGDIGLOQT-----QPYSIDTFKISGPICGMVIYARPD 67  
| : |  
OY 77 DGGDMFYT---GNPN--EYNTIPNSQSYSLLKSENSQITQIK---RYVCSNTSKLR 126  
| : |  
DB 68 TFTNSFYVTFYVNVGRASIIINYINTNMBENWSVLKAPGVVAEIRTPGNRLVFTSR--HIR 125  
| : |

```

Oy 127 TKSSAKYTTTSGKIVSTTQNSINSRRVYNAIDATNFPDDELRTKTRFFENQSYTSHK 186
Db 126 NLEDAQYISDYWLKXISISNAVLTENIP1-----TLNF-DQYDAGAAVAFVDRWFTQHP 180
Oy 187 SSTNSLYVHTWTIPRSLKIQNRMWEDYNGMTWAQSCYYKTGADGSGSESTRW--LAAQSI 244
Db 181 SD-----WASCVAK---EGLINSQNMGPHEMNH 207
Oy 245 PEPQNY--DGLWLDNDIALSGMAKSYNDVTGINQLSFTRIIIG--KGS-WYVINSGLD 298
Db 208 HMQGYLKGWV-----GIKEPGEETNNWMTINYLITYNIAHQRNQGSLGMYN----- 256
Oy 299 RGHAVITIIDQYGNKRYILFHAQYENSDEPLSSSI 332
Db 257 -----VSDGYSTTYKIL--NGENDQPHLRSTV 281

RESULT 4
571628
sensory transduction histidine kinase doka - slime mold (Dictyostelium discoideum)
C:Species: Dictyostelium discoideum
C:Date: 27-Nov-1997 #sequence_revision 12-Dec-1997 #ext_change 09-Jul-2004
C:Accession: S71628; S78068
R:Schuster, S.C.; Noegel, A.A.; Oehme, F.; Gerisch, G.; Simon, M.I.
EMBO J. 15, 3880-3889, 1996
A>Title: The hybrid histidine kinase doka is part of the osmotic response system of Dicty
A:Reference number: S71628; MUID:96324396; PMID:8670893
A:Accession: S71628
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-1670 <SCH>
A:Cross-references: UNIPROT:Q23901; UNIPARC:UPI000017CBFD; EMBL:X56869
A:Experimental source: strain AX2, substrain 214
R:Schuster, S.C.; Noegel, A.A.; Oehme, F.; Gerisch, G.; Simon, M.I.
submitted to the EMBL Data Library, March 1996
A:Description: The hybrid histidine kinase doka is part of the osmotic response system of
A:Reference number: S78068
A:Accession: S78068
A:Molecule type: DNA
A:Residues: 1-149,'E',151-219,'TRVLKLIQSTNNWYI',238-1671 <SCH>
A:Cross-references: UNIPARC:UPI000080237; EMBL:X56869; NID:91237201; PIDN:CAA65612.1; P
C:Genetics:
A:Gene: doka
C:Function:
A:Description: modulates cell response to changes in osmolarity; involved in spore forma
C:Keywords: phosphoprotein; signal transduction
F:1520-1629/Domain: response regulator homology #RRH2>
F:1568/Binding site: phosphate (Asp) (covalent) #status predicted

Query Match 6.5%; Score 115.5; DB 2; Length 1671;
Beet Local Similarity 22.3%; Pred. No.4.3; 91; Indels 99; Gaps 11;
Matches 64; Conservative 33; Mismatches

23 TGLDNCNIYANGIMIGVINITPTDEGNFVDIDVTIINDNIKIYDIDGSDIGSDGW 82
1347 TMLDLESSEKSTRENQDIIIV-----GSDKSTFVQ-----QILDVYTEKV-----I 1387
Oy 83 FTTGPNP-N-EYNTIPNSQSYSL-----LKSNSQITQIKRYVSCNT 122
Db 1388 LYGVPNPSKYN--ENSKVSYLVTPTTHSKLISILSKNMLKSNKNSFLYTTNNNTNNNT 1445
Oy 123 SKLRTKSPFAKYTTTSGKIVSTTQNSINSRRVYNAIDATNFPDDELRTKTRFFENQSY 182
Db 1446 NNIEKNS-----SIDSPILITSTSSIIITPILISNNLDLN----- 1481.
Oy 183 TSHKSTNSLYVHTWTIPRSLKIQNRMWEDYNGMTWAQSCYYKTGADGSGSESTRWLAAG 242
Db 1482 -NNNNNNNSILVS-----NNG-----GVDGNNVPSITLTI 1511
Oy 243 SIFPQNYGLMLNDIALSGMAHK-----SYNDVTGINQLSFTRIIIG 285
Db 1512 QQSQPKYTLVAEDNDINIKVVRQLEKGYTALVINGINKALEIIG 1558

```

## RESULT 5

glutamate 5-kinase [imported] - Streptococcus pneumoniae (strain TIGR4)

C/Species: Streptococcus pneumoniae  
C/Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 09-Jul-2004

C/Accession: P95107

R/Tettelein, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heidom, J.D.; Umeyam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzaple, neon, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001

A/Authors: Lottum, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison, A/Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.

A/Reference number: A95000; PMID:21357209; PMID:11463916

A/Accession: P95107

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-369 &lt;KUR&gt;

A/Cross-references: UNIPROT:Q97R95; UNIPARC:UPI0000051644; GB:AE005672; PIDN:AAK75055.1;

A/Experimental source: strain TIGR4

C/Genetics:

A/Gene: SP0931

C/Superfamily: glutamate 5-kinase

Query Match 6.2%; Score 111; DB 2; Length 369;

Best Local Similarity 25.0%; Pred. No. 1.2;

Matches 45; Conservative 34; Mismatches 73; Indels 28; Gaps 8;

31 YANGLMIGVITN--ITPTDEGNFVDIDVTLANIKIVYIDG-----SDIDG 78

113 YKHAQALSTLNRGAIPINENDSVVIDELKVGDNLTLSAQVAAMVQADLVFLTDVDG 172

79 SDGMFTGNPNENYNTIPNSQSYSLKSENSQITQIKRYVSCSN-TSRLRTKSPSAKYTTT 137

173 ----LYTGNP---NSDPRAKRLRIETIRNEIIDMAGAGSSNGTGMLTKIRAAATATE 225

138 SGKVISITONSINSRVINAIDATNFTDDE---LRTKETREPNOSYSHKSTSLVY 124

226 SGVPVVIC-SLKSMD---SMIEAETEDGSYFVAQEKGLRQKQMLAFVAGSQSISW 280

## RESULT 6

H97975

glutamate 5-kinase (EC 2.7.2.11) [imported] - Streptococcus pneumoniae (strain R6)

C/Species: Streptococcus pneumoniae

C/Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 09-Jul-2004

C/Accession: H97975

R/Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M, Y. P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001

A/Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.; A/Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A/Reference number: A97872; PMID:11544234

A/Accession: H97975

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-376 &lt;KUR&gt;

A/Cross-references: UNIPROT:Q8DQ61; UNIPARC:UPI0000083527; GB:AE007317; PIDN:AAK99636.1;

C/Genetics:

A/Gene: proB

C/Superfamily: glutamate 5-kinase

C/Keywords: phosphotransferase

Query Match 6.2%; Score 111; DB 2; Length 376;

Best Local Similarity 25.0%; Pred. No. 1.2;

Matches 45; Conservative 34; Mismatches 73; Indels 28; Gaps 8;

31 YANGLMIGVITN--ITPTDEGNFVDIDVTLANIKIVYIDG-----SDIDG 78

120 YKHAQALSTLNRGAIPINENDSVVIDELKVGDNLTLSAQVAAMVQADLVFLTDVDG 179

QY 79 SDGMFTGNPNENYNTIPNSQSYSLKSENSQITQIKRYVSCSN-TSRLRTKSPSAKYTTT 137

DB 180 ----LYTGNP---NSDPRAKRLRIETIRNEIIDMAGAGSSNGTGMLTKIRAAATATE 232

QY 138 SGKVISITONSINSRVINAIDATNFTDDE---LRTKETREPNOSYSHKSTSLVY 194

DB 233 SGVPVVIC-SLKSMD---SMIEAETEDGSYFVAQEKGLRQKQMLAFVAGSQSISW 287

## RESULT 7

B71605

hypothetical protein PFB0850C - malaria parasite (Plasmodium falciparum)

C/Species: Plasmodium falciparum

C/Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 21-Jul-2000

C/Accession: B71605

R/Gardner, M.J.; Tettelein, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.; Science 282, 1126-1132, 1998

A/Authors: M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.; A/Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.

A/Reference number: A71600; PMID:99021743; PMID:9804551

A/Accession: B71605

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-807 &lt;GAR&gt;

A/Cross-references: UNIPARC:UPI000017B60F; GB:AE001420; GB:AE001362; NID:G3845287; PIDN: A/Experimental source: clone 3D7

C/Genetics:

A/Gene: PFB0850C

Query Match 6.1%; Score 109; DB 2; Length 807;

Best Local Similarity 19.9%; Pred. No. 4.7;

Matches 68; Conservative 59; Mismatches 122; Indels 92; Gaps 13;

20 ILETG-LDNCNIYAN-----GLNMIGVITNITPTDEGNFVDID--DVTL 61

44 INSGHMSNDINIKNQEKKKKKKKKNTKRVININHTNIDKNNQODINKPEVIE 103

62 NNDIKIVYIDGSDIDGSDGMFTGNPNENYNTIPNSQSYSLKSENSQITQIKRYVSCSN 121

104 RDIINIKNDITNIIIDSS--YNEEGENNRNDINNINNINNINNINNINNINN--SCSN 156

122 TSRLRTKSPSAK-----VTTTSGK-----VISITONSINSRVINAIDA 161

157 NYGLKKKITLAKNDIKDESYNNENITTLNNKNNNNNNNNNNNNNNNNNNNN 216

162 TNFTDELRTKYS-----TRPNOSYSHKSTSLVYHTWTIPRSLKLNWRW 210

217 NNCSEKTELEQREKEYNKIRARIFSNFNKKQKVQKTEQNNLN-HTY-----LNNNII 268

211 EDVNNG---WTMAQSCYTKGADGSESTRMLAAGSIFPPGNTDGLMDNDIALSGAAHK 267

269 NNINNDNNOYAYINNRYH-----IYHNSNTHIYRNNIPICNINNH 310

268 SYVVDV-----GINOLFTRIIGKGFVVYISGLDRGH 301

311 APNIEKLNPPYTHDHIAVTNMYSTQKNNMKTKQIGH 351

## RESULT 8

T30857

glucosyltransferase - Streptococcus salivarius

C/Species: Streptococcus salivarius

C/Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 09-Jul-2004

C/Accession: T30857

R/Simpson, C.L.; Giffard, P.M.; Jacques, N.A.

Infect. Immun. 63, 609-621, 1995

A/Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for prin

A/Reference number: Z20909; PMID:95122197; PMID:7822030

A/Accession: T30857

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-1449 &lt;SIM&gt;

A/Cross-references: UNIPROT:Q55264; UNIPARC:UPI0000081668; EMBL:L35495; NID:G662378; PID



A:Molecule type: DNA  
A:Residues: 1-343 <KLE>  
A:Cross-references: UNIPROT:O51359; UNIPARC:UP10000057476; GB:AE001145; GB:AE00783; NIT  
A:Experimental source: strain B31  
C:Superfamily: Borrelia burgdorferi hypothetical protein BB0398

Query Match 6.0%; Score 107.5; DB 2; Length 343;  
Best Local Similarity 18.4%; Pred. No. 1.9;  
Matches 44; Conservative 36; Mismatches 80; Indels 77; Gaps 9;

QY 14 VSTFNVILETGLDNCNIYANGLMNIGVINITPTDDEGNFVDIDVTLANDNIKIYDIDG 73  
DB 141 INYKFLIMNGIDILN-----NEBKGNFYNALSLD---VQDY--- 177  
QY 74 SDDSGSCWFTYTGNNPNTYTP-----NSQSYSLKSNISQITQ 112  
DB 178 -----DSYFY---KKFLSI PRAHLKIDSRDYENVVTKINFFNPFVYRMGDIQD 229  
QY 113 IKRVSCSNSTR-----RTKFSFAKYTTSGKVISITONSINSRVVINAIDATNPTDD 167  
DB 230 VKRFVLSGNTSKLNTIDKNPFQSWDQKGGK-----SNSINTNSFLTMIIRLGRKN 284  
QY 168 ELRTYKTRFENOSYTSKSTNSLYVHTW-----TTPRSLKQNRWEDYNNGW 219  
DB 285 GIQFARKLLEADSSDDISYLESRGMDIHFWYFVKRIIVYPRD-----PEINNGWTW 335

## RESULT 12

H90053  
hypothetical protein fmb [imported] - Staphylococcus aureus (strain N315)

C:Species: Staphylococcus aureus  
C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004

C/Accession: H90053  
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc  
ma, A.; Mizutani, U.; Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;  
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratake, K.  
Lancet 357, 1225-1240, 2001  
A>Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A/Reference number: A89758; MUID:21311952; PMID:11418146  
A/Accession: H90053  
A/Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1038 <KUR>  
A:Cross-references: UNIPROT:Q99RD2; UNIPARC:UP100000CAB43; GB:BA000018; PID:G13702453; F  
C:Experimental source: strain N315  
C:Genetic:

A:Gene: fmb

Query Match 6.0%; Score 107.5; DB 2; Length 1038;  
Best Local Similarity 22.9%; Pred. No. 8.5;  
Matches 78; Conservative 39; Mismatches 130; Indels 93; Gaps 17;

QY 13 GVSTF-----NVILETG--LDNCNIYANGLMNIGVINITPTDDEGNFVDIDVT 61  
DB 249 GVSTARKVPEIKNGSVYMATGEILGNGNIRYTFNIEHKVEYANLEIMLFDPKTVQS 308  
QY 62 NNDIKIVDIYDSDSDSGFYTGNPNEVNTIPN--SQSYSLKSNISQITQIRYVSCS 120  
DB 309 NGBOKITSKLNGETEXTIPVY--NPGVNSYNNVNGSLETENKSNKCTHI--AYIKPM 365  
QY 121 NTSRLRTKSPSAKYTTTS-----GKVISITONSINSRVVINAIDATNPT 165  
DB 366 NNGOSNTVSTGTLTSSNLAGGQPYKRYEYLCKDELQOS-----VYANTSDTNKKF 419  
QY 166 DDELRTTKETR-----FENOSYTSKSTNSLYV--HTWTIPRSLKQNRWEDYN--- 214  
DB 420 D-----VTKENAGKLSVDNGSYSLMDKLDKTYVIHTGELYQSSDQVNFRTELYGYPER 475  
QY 215 -----NGW--TWAGS--CYKGTGADGSESTRWLAAGSFPFGANTDGLMDNDI--- 259  
DB 476 AYKSYVYVGGRLTWNDGLVLYSNKADGNGKN-----GGIIO-----DNDPEYKE 520  
QY 260 -----ALSGMAHKSYNVDTGINQLS-----FTRIIGKG 287

DB 521 DTAKGTMSGYDAKQIIFTEFNQDNTPLDIDYHTALDGE 560

## RESULT 13

A86827

hypothetical protein ydfg [imported] - Lactococcus lactis subsp. lactis (strain IL1403)

C:Species: Lactococcus lactis subsp. lactis  
C>Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 09-Jul-2004

C/Accession: A86827  
R:Bellocin, A.; Winkler, P.; Mauger, S.; Jallion, O.; Malarme, K.; Weissenbach, J.; Ehrlich  
Genome Res. 11, 731-753, 2001  
A>Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp.

A/Reference number: A86825; MUID:2125106; PMID:11337471

A/Accession: A86827  
A/Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1072 <STO>  
A:Cross-references: UNIPROT:Q9CF64; UNIPARC:UP100000D4491; GB:AE005176; PID:G12724625; P  
C:Experimental source: strain IL1403  
C:Genetic:

A:Gene: ydfg

Query Match 6.0%; Score 107; DB 2; Length 1072;  
Best Local Similarity 21.2%; Pred. No. 9.7;  
Matches 65; Conservative 52; Mismatches 136; Indels 54; Gaps 10;

QY 10 DWTGVSTFNVILETGLDNCNIYANGLMNMT--GVINITPTD-----EGNFVDID 57  
DB 610 DSEISTGAFSLNQTSSPSSMSINSPLSLTSSRSATQNSSEATKYDNN 669  
QY 58 DVTLANDNIKIYDIDSDIDG---SPGWFYTGNNPNTIPNOSYSLKSNISQITQIR 114  
DB 670 SSTHSSNITLNSGNDSDSDSDSSNLSNPN---LETQTLSSKPEVNNISENP 725  
QY 115 RYVSCSN-----SLRTKFSFAKYTTSGKVISITQ--NSINSRVVINAIDATN 163  
DB 726 KKVSSNSVQENSTDEHMTNPPKSSISSPSTTSSSQKESQSNLNTTGGINNPIFPNN 785  
QY 164 FTDELRTTKETRFENOSYTSKSTNSLYVHTW-----TWTIPRSLKQNRWED 212  
DB 786 SSSSENSAASITLTSYNNN--SESSFTGCLYISNDAQRDNGSEISHLPSNSNEN----- 838  
QY 213 YNNGWTWAGSCYKGTGADGSESTRWLAAGSIFPPGYDGLMDNDIALSGMAHKSYNVD 272  
DB 839 -NVSSTQQAITLESSKSTNKRSSLSLITNSTHPQNE-----DNQSNSSDEVKSNNV 892  
QY 273 TGINQLS 279  
DB 893 SLTGQLN 899

## RESULT 14

JH0284

125K surface antigen M17 precursor - Entamoeba histolytica

C:Species: Entamoeba histolytica  
C>Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 09-Jul-2004

C/Accession: JH0284  
R:Edman, U.; Meraz, M.A.; Raueser, S.; Agabian, N.; Meza, I.  
J. Exp. Med. 172, 879-888, 1990

A>Title: Characterization of an immuno-dominant variable surface antigen from pathogenic

A/Reference number: JH0284; MUID:90354789; PMID:1696956

A/Accession: JH0284  
A/Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1114 <EDM>

A:Cross-references: UNIPROT:Q05352; UNIPARC:UP1000007AD25  
A:Experimental source: strain HMI IMSS

Query Match 6.0%; Score 107; DB 2; Length 1114;  
Best Local Similarity 21.8%; Pred. No. 10;  
Matches 77; Conservative 37; Mismatches 125; Indels 114; Gaps 19;

QY 21 LETGLDNCNIYANG--LNMIGVIINPTPTDEGNFVDIDV--TLNDNIKIYDIYDGS 75  
::||::||::||::||::||::||::||::||::||::||::||::||::||::||  
Db 187 IQVGIGCKCNPNPDSQMLANVMNMSNDRMFIDSIGEDGLNTQYIINDTFTKI-----GSP 241  
QY 76 IDG-----SD-----GWFYTNPN-----EYNTIPNSQSYLSKSENSQITQIKRYVSC 119  
242 FGGWIIYLRSDTFTNSFYVTFNSVGRAPITINYNTTNEWNSVLRNAPGVAEIR----- 296  
QY 120 SNTSRLRTKSPSAKVTTTSGKVISITONSINS-----SRVYNALDANFTDDELRTT 172  
Db 297 -----TPGNRLVLTISRNIRSLDAQYISDFWLKALISINYA----- 332  
QY 173 KETRFENQSYTSH-----KSTNSLYVHTWTIPRSLKLONMWEDYNNGWMTAOSCYKKT 227  
Db 333 --VTLENIPITLNFDPQVDAGAAVAYGRWFT-----QNPSDMAAC---V 373  
QY 228 GADGSESTRW--LAAGSIFPPGNY--DGLMLNDIALSGMAKSYVNDTGINQLSFTRI 283  
Db 374 GKDGILNYGWMGPIHEMNHHMQGYLKGHW-----GISNPGERTNNVMTSINYIYTYNI 428  
QY 284 IG---KGFS--WVYNISGLDRGHAVIIDIYGNKYRILFHAGYENS DPLYSSSI 332  
Db 429 AGHRNQLSGMNY-----VSDGYSTTYKIL--KGENDQPHLSYV 466

## RESULT 15

B81963  
IGA-specific serine endopeptidase (EC 3.4.21.72) MMA0457 [imported] - Neisseria meningitidis  
C/Species: Neisseria meningitidis  
C/Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 09-Jul-2004  
C/Accession: B81963  
R/Parikhil, J.; Achman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel,  
Nature 404, 502-506, 2000  
A/Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.  
A/Reference number: A81775; MUID:20222556; PMID:10761919  
A/Accession: B81963  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1,1449 <PAR>  
A/Cross-references: UNIPROT:Q9TMB4; UNIPARC:UPI00000C4995; GB:AL162753; GB:AL157959; NID  
A/Experimental source: serogroup A, strain Z2491  
C/Genetics:  
A/Gene: iga2; MMA0457  
C/Superfamily: IGA-specific metalloendopeptidase  
C/Keywords: hydrolase; serine proteinase

Query Match 6.0%; Score 107; DB 2; Length 1449;

Best Local Similarity 21.1%; Pred. No. 15;  
Matches 71; Conservative 46; Mismatches 126; Indels 94; Gaps 17;

QY 9 SDWTVGSTFNVILFTGLDNCNIYANGLMIGVIINPTPTDEGNFVDIDV-----V 59  
|||::||::||::||::||::||::||::||::||::||::||::||::||::||  
Db 722 SDWTVGLT--NCVEKRTITD-----KVIALFTYTDISGN-VSLADHAHLNLGLA 767  
QY 60 TLNDNIKIYDIYDSDIDSDGWFYTGPNFYNTIPN--SQSYLSKSENSQITQIKRYVS 118  
|||::||::||::||::||::||::||::||::||::||::||::||::||::||  
Db 768 TLNGNLS-----ANGDTRYVSHNATONGNLSIVGNAQAT--FNQATL 808  
QY 119 CSNTSRLRTKSPSAKVTTTSGKVISITONS--INSRVYNAIDAFTDDELRTTKETRF 177  
|||::||::||::||::||::||::||::||::||::||::||::||::||::||  
Db 809 NGNTSASGNASFNLSNNAQNGSLTLDNAKANVSHSALNG-----NVSADKAVFHF 861  
QY 178 ENQSYTSHKSTNSLYVH---TWIIPRSLKLONMWEDYNNGWMTAOSCYV--KTGADG 231  
|||::||::||::||::||::||::||::||::||::||::||::||::||::||  
Db 862 ENSRFTGQLSGSKDTALHKDSEWTLPSGTEGLN---NLDNATITLNSAYRHDAAGQOT 918  
QY 232 GS-----ESTRWLAAGSIFPPGNYDGLMLNDIALSGMAKSYVNDTGINQLSFTRI 284  
|||::||::||::||::||::||::||::||::||::||::||::||::||::||  
Db 919 GSVSDTTPRRRRRLRL--SVTPPTSVESRF--NLTIVNGK-----LNGGTRFRFM 964  
QY 285 GKGFWSM-----VYNISGLDRGHAVIIDIQ 308  
::||::||::||::||::||::||::||::||::||::||::||::||::||

Db 965 SELFGRSDKLKLASSSEGTYYTLAVNNTGNBPVSLDQ 1001

Search completed: January 30, 2006, 09:49:39  
Job time: 7.82461 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: January 30, 2006, 09:40:48 ; Search time 26.6268 Seconds  
(without alignments)  
8849.971 Million cell updates/sec

Title: US-09-889-874A-22

Profile score: 1784

Sequence: 1 FTLRSDMSDMVTGFSTFNV.....ILFHAGYNSDPYLSSTIVY 334

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: uniprot\_05.80:\*

2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1749	98.0	327	2	Q9EVR8_XENBV
2	145.5	8.2	340	2	Q7NAU4_PHOIL
3	136	7.6	334	2	O52884_COXBU
4	129.5	7.3	1550	2	Q4JL53_LACRE
5	126	7.1	892	2	Q5HR16_STAEQ
6	124.5	7.1	931	2	Q9K113_STAPQ
7	124.5	7.0	293	2	Q4N914_THBPA
8	124	7.0	1056	2	O8C072_STARP
9	124	7.0	1092	2	Q70022_STARP
10	122.5	6.9	404	2	Q93109_VIBR
11	122.5	6.9	990	2	Q97VW0_SULSO
12	121	6.8	1765	2	Q7V885_PROKM
13	119	6.7	814	2	O52984_ECOLI
14	119	6.7	883	2	O832P7_ENTRA
15	119	6.7	1361	2	Q4L555_STAHJ
16	118.5	6.6	537	2	Q54WL1_DICDI
17	118	6.6	773	2	Q7UCU4_SHIFL
18	118	6.6	1502	2	Q7R8L6_PLAYO
19	118	6.6	1502	2	Q7R8L6_PLAYO
20	118	6.6	6761	2	O6L8X8_PLAF7
21	117.5	6.6	627	2	Q7BD11_PLAYO
22	117.5	6.6	640	1	APRN_ENTHI
23	117.5	6.6	3779	2	Q5GQB9_9CAUD
24	117	6.6	941	2	Q7R1W5_PLAYO
25	116.5	6.5	1054	2	Q9LAB9_PIRCK
26	116.5	6.5	1092	2	O8EUS1_MYCPE
27	115.5	6.5	861	2	O8PVN5_METMA
28	115.5	6.5	962	2	Q5S854_DICDI
29	115.5	6.5	1670	2	Q23901_DICDI
30	115.5	6.5	1671	2	O86985_DICDI
31	115.5	6.5	2835	2	O8G9Q2_LEUME

32	115.5	6.5	3623	2	Q7R069_PLAYO	Q7R69 plasmodium
33	115	6.4	672	2	O6FK03_CANGA	O6FK03 candida gla
34	115	6.4	2001	2	O6KCM6_ECOLI	O6KCM6 escherichia
35	115	6.4	2387	2	O8FFP8_ECOLI	O8FFP8 escherichia
36	114.5	6.4	456	2	O6CUX7_KLULA	O6CUX7 kluyveromyc
37	114.5	6.4	1392	2	Q4QNU0_HAB18	Q4QNU0 haemophilus
38	114.5	6.4	3078	2	O8IKH4_PLAF7	O8IKH4 plasmodium
39	114	6.4	5251	2	O81RD4_PLAF7	O81RD4 plasmodium
40	113.5	6.4	878	2	O6B196_DEBHA	O6B196 debaryomyce
41	113.5	6.4	1054	2	Q9LAF1_RICK	Q9LAF1 rickettsia
42	113	6.3	412	2	O54X20_DICDI	O54X20 dictyostell
43	113	6.3	584	2	Q9KH13_MYCBA	Q9KH13 mycoplasma
44	113	6.3	2110	2	O8IBF2_PLAF7	O8IBF2 plasmodium
45	112.5	6.3	608	2	Q4KH09_PLACH	Q4KH09 plasmodium

## ALIGNMENTS

### RESULT 1

Q9EVR8\_XENBV  
AC Q9EVR8\_XENBV PRELIMINARY; PRT; 327 AA.  
ID Q9EVR8;  
DT 01-MAR-2001 (TREMREL. 16, Created)  
DT 01-MAR-2001 (TREMREL. 16, Last sequence update)  
DT 01-MAR-2001 (TREMREL. 16, Last annotation update)  
DE Nematocidal protein.  
GN Name=xmpl;  
OS Xenorhabdus bovienii.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Xenorhabdus.  
OX NCBI\_Taxid=40576;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=173;  
RA Morgan J.A.W., Quisley M., Ellis D., Jarrett P.;  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AJ296651; CAC19492.1; -; Genomic DNA.  
SQ SEQUENCE 327 AA; 36684 MW; 8417B96FB70D60D CRC64;

QY	8	MSDWTGVSTFNVLTETGDNCTNYANGLMVGIYITPTDDEGNVYDIDVTIANIKI	67
DB	1	MSDWTGVSTFNVLTETGDNCTNYANGLMVGIYITPTDDEGNVYDIDVTIANIKI	60
QY	68	VDYIDSDIDSGDWGYTGNPNRYNTIPNSQSYSLKSENSQITQIKRYVCSNTSLRT	127
DB	61	VDYIDSDIDSGDWGYTGNPNRYNTIPNSQSYSLKSENSQITQIKRYVCSNTSLRT	120
QY	128	KSPFAVYTTSGVVISITONSINSRVINAIDATFTDELATTKETRENSQYSHKS	187
DB	121	KSPFAVYTTSGVVISITONSINSRVINAIDATFTDELATTKETRENSQYSHKS	180
QY	188	STNSLYVHTTTPRSIKLQWRWEDYNNGTWAGSCYKTAGDGSESTRMLAAGSIFPP	247
DB	181	STNSLYVHTTTPRSIKLQWRWEDYNNGTWAGSCYKTAGDGSESTRMLAAGSIFPP	240
QY	248	GNTDGLMLNDLALSGMAHKSYYVDGINOLSTRTIIGKFSVWYNSIGLDRHAYIID	307
DB	241	GNTDGLMLNDLALSGMAHKSYYVDGINOLSTRTIIGKFSVWYNSIGLDRHAYIID	300
QY	308	QYGNKTRIIIFHAGYNSDPYLSSTIVY 334	
DB	301	QYGNKTRIIIFHAGYNSDPYLSSTIVY 327	
RESULT 2			
QY	Q7NAU4_PHOIL		
ID	Q7NAU4_PHOIL PRELIMINARY; PRT; 340 AA.		
AC	Q7NAU4_		

DT 01-MAR-2004 (Tremblrel. 26, Created)  
 DT 01-MAR-2004 (Tremblrel. 26, Last sequence update)  
 DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)  
 DE Unknown protein.  
 GN OrderedLocustNames=plu2221;  
 OS Photobacterium lumenescens (subsp. laumondii).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Photobacterium.  
 OX NCBI\_TaxID=141679;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=TT01;  
 RX MEDLINE=D295627; PubMed=14528314; DOI=10.1038/nbt886;  
 RA Duchaud E., Rustiock C., Frangeul L., Buchrieser C., Givaudan A.,  
 RA Taouit S., Boco S., Bouraux-Rude C., Chandler M., Charles J.-F.,  
 RA Daba B., Desros R., Derzelle S., Freysinet G., Gaudreau S.,  
 RA Meligne C., Lanois A., Powell K., Siguler P., Vincent R., Wingate V.,  
 RA Zouine M., Glaeser P., Boemare N., Danchin A., Kunst F.,  
 RT "The genome sequence of the entomopathogenic bacterium Photobacterium  
 lumenescens";  
 RL Nat. Biotechnol. 21,1307-1313(2003).  
 DR EMBL, BX571866; CAI4514.1; -; Genomic\_DNA.  
 DR Photobact; plu2221; -;  
 KW Complete proteome.  
 SQ SEQUENCE 340 AA; 38470 MW; 818864919F82175E CRC64;  
  
 Query Match 8.2%; Score 145.5; DB 2; Length 340;  
 Best Local Similarity 22.9%; Pred. No. 0.033;  
 Matches 74; Conservative 59; Mismatches 137; Indels 53; Gaps 15;  
  
 QY 29 NIYANGIMIGVIINITPTD-DEGNFVIDDVTINDNK-----IYDIIDGSDIDGSD 80  
 DB 20 NFYANGHQCMKSVLKQEKYKGDWVCL--ALSDAKRSIQVAALSDSLIYQOLKMS 76  
 QY 81 GMPYTGNEPNEP-----NTIPNSQSYS-----LKSNSQ--ITQIKRVYSC-SNT 122  
 DB 77 GMTTIDANKKTDGLLNGVYSGADIFTEBPVIRAGDCCCTNENYQNSVKSPEIICYVSS 136  
 QY 123 SRLTKSSAKVT--TISGKVIISITONSIN-----SSRVINAIIDATNTDELATTKETR 176  
 DB 137 NRTSTELYMAKWTBEDTNGKRTLTNNMSVGDEVPDSKLLALAPYAINAQLHSTNITL 196  
 QY 177 FENOSYTHSKSTN--SLYVHTWTIPRLKLOMMRWEDYNNNGMTWAGSCYKKGADGSE 234  
 DB 197 FDKTEEPF-KSDTHHQITINIRKWLPIRLLEGDRKRVNNIYVLGKS-----SS 245  
 QY 235 STRWLAAGSIFPPGNYDGLMLDNDIALSGMAHKSYNVDTGINQISFTRIICKGFSWVNI 294  
 DB 246 DDFLTLTRARVFRGT-SYVNAARNDSGGCVMDSYDVTPTQLAAVLAHLVYGCSTTGY 304  
 QY 295 SGLDRGHAIVIIIDYG--NKYRI 315  
 DB 305 --VDGYHDVTIIDNYGCOHKFRI 325  
  
 RESULT 3  
 OS2884 COXBU  
 ID OS2884 COXBU PRELIMINARY; PRT; 334 AA.  
 AC OS2884;  
 DT 01-JUN-1998 (Tremblrel. 06, Created)  
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE Hypothetical protein ort 334b.  
 GN Name=ort 334b;  
 OS Coxiella burnetii.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;  
 OC Coxiellaceae; Coxiella.  
 OX NCBI\_TaxID=777;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Lauenschlaeger S., Jaeger C., Willems H., Baljer G.,  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL, Y15898; CAA75845.1; -; Genomic\_DNA.  
 KW Hypothetical protein; Plasmid.  
 SQ SEQUENCE 334 AA; 39048 MW; 6BDC7E279851DCC CRC64;  
  
 Query Match 7.6%; Score 136; DB 2; Length 334;  
 Best Local Similarity 21.3%; Pred. No. 0.15;  
 Matches 77; Conservative 49; Mismatches 155; Indels 80; Gaps 17;  
  
 QY 11 WTGVSTFNVILTE-----GLDNCNIYANGLMIGVIINITPTDEGNFVIDDVTIANNIK 66  
 DB 3 WENPFRSVIKLAPPGHPITQANIIYANGRNQVIRVLRLDREGNPLKLTIEDLREKCS 62  
 QY 67 IVDYIDGSDID-----GSDGMPYTGPNEX-----NTIPNSQSYS-LKSNSQ 109  
 DB 63 LIYQYEEBISYKSGTESANISQNMWYGTAVNDYQCNFPLVP-----YSLNTNDSDDK 118  
 QY 110 ITQIKRVYSCNTRSLTKTSFSAKVTTSKVIISITONSINSGRVINAIIDATNFTDEL 169  
 DB 119 VSYIDYTL-----KSGERTGKHIGLCININGENYKLSGKFKGVVETIAPKKYTDNI 174  
 QY 170 RTTKETREFNQSYTHSKSTNSLYHTWTIPRLKLOMMRWED-----YNNGTWA 220  
 DB 175 TFRRESK--KIGTTNWYGAQGYNVEKTYLTIKDSFRIKKGIDINHVFRYGDWK 232  
 QY 221 QSCYKKGADGSGSESTRWLAAGSIFPPGNYDGLMLDND-IALSGMAHKSYN----- 270  
 DB 233 FGFFLDRSGD-----KRTGQHLVPIEFS--LRIDKVFSGKHYQHNVFFYEDAMC 281  
 QY 271 --VDTGINQISFTRIICKGSWYNNISGLDRG-----AVIIDIQYEN--KIRLFLNAGY 321  
 DB 282 FFWVTGV--FTVPIGQ--WTY--GNEKCNEMKCFYFYDQYNGEKFTRICDDGW 331  
 QY 322 E 322  
 DB 332 E 332  
  
 RESULT 4  
 OS04UG3 LACRE  
 ID 04UG3 LACRE PRELIMINARY; PRT; 1550 AA.  
 AC 04UG3;  
 DT 13-SEP-2005 (Tremblrel. 31, Created)  
 DT 13-SEP-2005 (Tremblrel. 31, Last sequence update)  
 DT 13-SEP-2005 (Tremblrel. 31, Last annotation update)  
 DE Hypothetical protein.  
 GN ORNames=lr1612;  
 OS Lactobacillus reuteri.  
 OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;  
 OC Lactobacillu.  
 OX NCBI\_TaxID=1598;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=ATCC 55730;  
 RA Wall T., Bach K., Roos S., Jonsson H.,  
 RT "Bioinformatic identification of the secretome of Lactobacillus  
 reuteri ATCC 55730";  
 RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, DQ074912; AY86887.1; -; Genomic\_DNA.  
 KW Hypothetical protein.  
 SQ SEQUENCE 1550 AA; 172371 MW; 9A7052287A9014D7 CRC64;  
  
 Query Match 7.3%; Score 129.5; DB 2; Length 1550;  
 Best Local Similarity 23.7%; Pred. No. 2.8;  
 Matches 76; Conservative 55; Mismatches 139; Indels 51; Gaps 16;  
  
 QY 39 GVIINITPTDEGNFVDI-DDVTIND-----NIKIYDIIDGSDIDGSDGW 82  
 DB 69 GIATPDPASAEBSVQSGSTYVNDQTSFQVDSNIYGTAVNVKMDYENNSDIYNLTG- 127  
 QY 83 FYTGPNPENVNTIPNSQSYSLKSENSQITQIKRVYSCNTRSLTKTSASAVTTTSG--- 139  
 DB 128 -YHGDTRHIVPMLNDFNDAGIIGQTTSG--ISSITLTKNLVQSIAFAISKISGYN 184



```

QY 140 -KYISTON---SINSRVINA---IDATFTGDEL-----RTKETRENOSSYSHK 186
Db 185 NKVIA-TDNDMSVFPASSGGLVNAADLSNLTSTNTNMKYLPAYNKGLKTVGLISEMDTGHV 243
QY 187 S8TNSLVHTWTTIPRLSKLQNNRWEVDYNNGWMTAOSCYTKGADGSESTFWLAAGSIFP 246
Db 244 SDMASHFEYSALQSGAGNIGKMNNTNVTYSMMNMFPESDVSTKRTIGHNNTSKVVDMGIMF- 302
QY 247 PGNYDGLM--LDN-DIALSGMAHKSYNDVTGINOLSPFRIITGKGSW--VYNISGLDRGHA 302
Db 303 -HSYKCEWGLDSNWDVSNVKNMVEFVSSDRLGELSPNGNITK---WKIKGNANLVR--- 355
QY 303 VIIIIDGNKRYRLFLFAGYEN 323
Db 356 MFSFDRSTSPYRVV--ASNEN 374

```

RESULT 5	OSHR16_STAE0	
ID	OSHR16_STAE0 PRELIMINARY;	PRT; 892 AA.
AC	OSHR16;	
DT	10-MAY-2005 (TrEMBLrel. 30, Created)	
DT	10-MAY-2005 (TrEMBLrel. 30, Last sequence update)	
DT	10-MAY-2005 (TrEMBLrel. 30, Last annotation update)	
DE	Sarg protein.	
GN	OrderedListNames=SERP0207;	
OS	Staphylococcus epidermidis (strain ATCC 35984 / RP62A).	
OC	Bacteria; Firmicutes; Bacillales; Staphylococcus.	
OX	NCBI_TaxID=176279;	
	[1]	
RP	NITELIOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA]	
RA	PubMed=15774886; DOI=10.1128/JB.187.7.2426-2438.2005;	
RA	Gall S.R., Fouts D.E., Archer G.L., Mongodin E.F., Deboy R.T.,	
RA	Ravel J.P., Paulsen I.T., Kolonay J.F., Brinkac L.M., Beaman M.J.,	
RA	Adelson R.J., Daugherty S.C., Madupu R., Angioli S.V., Durkin A.S.,	
RA	Haft D.H., Vamathevan V.J., Khouri H., Uitterlbeck T.R., Lee C.,	
RA	Dimitrov G., Jiang L., Qin H., Weidman J., Tran K., Kang K.H.,	
RA	Hance I.R., Nelson K.E., Fraser C.M.;	
RT	"Insights on evolution of virulence and resistance from the complete	
RT	genome analysis of an early methicillin-resistant <i>Staphylococcus</i>	
RT	<i>aureus</i> strain and a biofilm-producing methicillin-resistant	
RT	<i>Staphylococcus epidermidis</i> strain."	
RL	J. Bacteriol. 187:2426-2438(2005).	
DR	EMBL; CP000029; AAW5357L.1; -; Genomc_DNA.	
DR	SMR; OSHR16; 237-557.	
DR	TIGR; SERP0207; -.	
DR	GO; GO:0009386; C:cell surface; IEA.	
DR	GO; GO:0005618; C:cell wall; IEA.	
DR	InterPro; IPR008454; Cna_B.	
DR	InterPro; IPR001899; Girm_pos_anchor.	
DR	Pfam; PF05738; Cna_B_2.	
DR	Pfam; PF00746; Girm_pos_anchor.1.	
DR	TIGRFAMs; TIGR01167; LPTXG_anchor.1.	
DR	PROSITE; PS00647; GRAM_POS_ANCHORING.1.	
KW	Cell wall; Complete proteome.	
QO	SEQUENCE 892 AA; 9874 MW; EC3C932B58BF497 CRC64;	

```

Query Match      7.1%; Score 126; DB 2; Length 892;
Best Local Similarity 21.0%; Pred. No. 2.5;
Matches 78; Conservative 51; Mismatches 141; Indels 102; Gaps 16

QY          9 SDMTGVSTFVNLLEGLDNCNIYANGLMIG--VIINITPTDEGNFYDIDVT----- 60
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db         398 SMPFNIDTKHTVEQ-----TIYNPLAKYSAKETWNVNISGNDBESTI-IDSTIIIKYYK 451

QY        61 -----LNDNKIVDYIDSGSIDSGMFPYTGPNPEYNT-----IPNS 97
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db        452 VGDNQNLPEDSNRIVDYSEYEDVT-NDVYAQLGNNDNVNINFGRIDSPYIIKVISKDPKK 510

QY        98 QSYSLSKSENSQIQIOIKRKYVSCSNTSLRLTSSFSKAV---TTTSG-----KVIST- 144
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db       511 DDYTTIIQQTVMMQTINEY----TGEPRTIASYDTIAFTSSGGGOGDLPEPKRYKKGID 565

```

```

Qy 145 -----TQNSINSSVVVNAIDANFPPDEDLRTTKERFENQSVYTHKSSTSL-Y 123
Db 566 YWEDVDKDGQIQNTYNDNEKPLSNVLVTLTPDG--TSKSVRTDEBKQYFPDGKGLTY 622
Qy 194 VHTWTIPRLAKLQNRMEDY-----NNGWTMAQSCYCYTKGADGSEST 236
Db 623 KITFFP-----EGYTPPLKHSSTNALDSEGNSSVVTITNGQDDMTIDSGPQTR 672
Qy 237 RMLAAGSIFPPGNVYDGLMLNDIALSGNAHKSINVYDVGINGQLSFTRIIGKGFSSVYNISG 296
Db 673 KYSLSGNYVWYDITNKKQIQGDDEKISGV-KVTLKDENGNIISPTTIDENG--KIQPDN 727
Qy 297 LDRGHAVITIIDQ 308
Db 728 LNSGNYVAFDK 739

```

Q9K113	STAEF	RESULT 6
ID	Q9K113_STAEF PRELIMINARY;	PRT; 931 AA.
AC	Q9K113;	
DT	01-OCT-2000 (TrEMBLrel. 15, Created)	
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)	
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)	
DE	Fibrinogen-binding protein SdrG.	
GN	Name=sdrG;	
OS	Staphylococcus epidermidis.	
OC	Bacteria; Firmicutes; Bacillales; Staphylococcus.	
OX	NCBI_TaxID=1282;	
LN	[1]	
RP	NOCLROTIDE SEQUENCE.	
RC	STRAIN=K28;	
RX	MEDLINE=20340957; PubMed=10878118;	
RA	McCrea K.W., Hartford O., Davis S., Ni Eidhin D., Lina G.,	
RA	Speziale P., Foster T.J., Hook M.;	
RT	"The serine-aspartate repeat (Sdr) protein family in Staphylococcus	
RT	epidermidis.";	
RL	Microbiology 146:1535-1546(2000).	
EMBL	AF245042; AAF72510.1; -; Genomic_DNA.	
PDB	1R17; X-ray; A/B=274-598.	
PDB	1R19; X-ray; A/B/C/D=274-598.	
DR	GO; GO:0009986; C:cell surface; IEA.	
DR	GO; GO:0005618; C:cell wall; IEA.	
DR	GO; GO:0016020; C:membrane; IEA.	
DR	InterPro; IPR008454; Cna_B.	
DR	InterPro; IPR005877; Gpos_YsIRK.	
DR	InterPro; IPR001899; Gram_pos_anchor.	
DR	Pfam; PF05738; Cna_B_2.	
DR	Pfam; PF00746; Gram_pos_anchor_1.	
DR	Pfam; PF04650; YsIRK_signal_1.	
DR	TIGRFAMs; TIGR01167; LpYTG_anchor_1.	
DR	TIGRFAMs; TIGR01168; YsIRK_signal_1.	
DR	PROSITE; PS50847; GRAM_POS_ANCHORING; 1.	
CM	Cell wall.	
Q9	SEQUENCE 931 AA; 102955 MW; 591857097027116 CRC64;	

```

Query March 7.1k; Score 126; DB 2; Length 931;
Best Local Similarity 21.0k; Pred. No. 2.7;
Matches 78; Conservative 51; Mismatches 141; Indels 102; Gaps 16;

QY 9 SDMTGVTSTFVILLETGDNLCNIVANGLMIG--VIINITPTDDEGNFVDDVDT----- 60
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 437 SMFNIDTIDTKHHVWQ-----TIYNPLRYSAKETNNVMSGDGGSTI-IDDTIIKYYK 490

QY 61 -----LNDNIKIYDYIDGSDIDGSDGWFYTGPNENYNT-----IPNS 97
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 491 VGDNONLPDSENRIYDSEYEDVT-NDRYAOLGNNDNVINFGNIDSPYIIKIVISKYDPNK 549

QY 98 QSVSLKSENSQITQIKRYVSCNTSLRLRKSPAKV--FTTSG-----KVSI-- 144
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 550 DDYTTIQVTVMOTTINBY-----TGEFRFASDYNTITAFSTISSQGGGDLPEETVYKIGD 604

QY 145 -----TONSINSSRVVINALDAITNFTDELRTTKETRFENQSVYSHKSSSTNSL-Y 193

```

[illegible]

```

Oy 147 INISSSVAVNAIDATNFTDDELTRTKEFENQSYSHSSTSLVYVHTMTI PRSLKQ 206
Db 169 NSVNTS-----VNSVAV-----TSVNSSVNT-----SVNLL 193
Oy 207 NMRREDYNNNGWTMAQSCYYK-TGADGGSSESTRMLAAGSIFPPGNVYG-----LWLDN 257
Db 194 GNDMSYDR-----VYDLGVIVKVSGGDGDGGSFRRKHV-PLGGANGANGPGGDVYIEC 248
Oy 258 DIALSGM-----AKSVYNDVGINOLSTRILIGK 287
Db 249 NDSVSDLRWFKANKIYKXANGNN-----GKG 274

RESULT 8
O8C072_STAEP
ID O8C072_STAEP PRELIMINARY; PRT; 1056 AA.
AC O8C072;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Ser-Asp rich fibrinogen-binding, bone sialoprotein-binding
protein.
GN OrderedLocusNames=SE0331;
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
[1]
[1] NUCLEOTIDE SEQUENCE (LARGE SCALE GENOMIC DNA).
RP STRAIN=ATCC 12228;
RC PubMed=12950922; DOI=10.1046/j.1365-2958.2003.03671.x;
RX Zhang Y.-O., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,
RA Qin Z.-Q., Mao Y.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z.,
RA Yuan Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.;
RT "Genome-based analysis of virulence genes in a non-biofilm-forming
RT Staphylococcus epidermidis strain (ATCC 12228).";
RL Mol. Microbiol. 49:1577-1593 (2003).
DR EMBL; AE016745; AAC03928.1; -; Genomic_DNA.
DR HSBP; 053653; IN67.
DR SMR; O8C072; 278-598.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR InterPro; IPR008454; Cna_B.
DR InterPro; IPR005877; Gpos_Ysirk.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003367; tap_3.
DR Pfam; PF05738; Cna_B_2.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02412; TSP_3; 1.
DR Pfam; PF04650; Ysirk_signal; 1.
DR TIGRFAMs; TIGR01167; LpxTG_anchor; 1.
DR TIGRFAMs; TIGR01168; Ysirk_signal; 1.
DR PROSITE; P850847; GRAM_POS_ANCHORING; 1.
KW Cell wall, Complete proteome.
SQ
SEQUENCE 1056 AA; 115728 MW; 015869A9E5CA2723 CRC64;

Query Match 7.0%; Score 124; DB 2; Length 1056;
Best Local Similarity 21.0%; Pred. No. 4.3;
Matches 78; Conservative 50; Mismatches 142; Indels 102; Gaps 16

Oy 9 SDMTGSVFNVILETGLDNCNIYANGLMIG-VIINTPTDDEGNFVDDVT----- 60
Db 439 SMFTWIDTKHNTVEQ-----TIYINPLRYSAKETVNVISGNGDEGST-IDDSTIIKYK 492
Oy 61 -----LNDNKIYVYIDGSDIDSGDGFYGNPNREVT-----IPNS 97
Db 493 VGDNONLDPDSNRIYYSSEYEDVT-NDDYAQLGNNDVAINFGNIDSPYIIKVISKIDPK 551
Oy 98 OGSYLKSENQITQIKRYVCSNTRISRTKFSFAKV-TTTSG-----KVYSI-- 144
Db 552 DDTYITIGQVIMQITINEX-----TGEKRTISYDNTIAFSSSSGGGGDLDPPEKTYKLG 606

```

```

Qy 145 -----TONGSRRVAINADATNFTDELRTTKETRFENQSYTSHKSTNSL-Y 193
Db 607 YWEDVDKQIGQNTDNEKRLSNVLTLYTPDG---TSKSVRTDEBQKQPDGLKNGLT 663
Qy 194 VHTWTIPRSIKLQNMWEDY-----NNGWTAAQSCYRTGADGSEST 236
Db 664 KTFETP-----EGYPTLKHSQTNPALDSEGNVWVTINQDDMTIDSGFYQTP 713
Qy 237 RMLAAGSIFPPGNYDGLMNDIALSGMAHKSINVDGINQLSFTRIGGFSMVNNSG 296
Db 714 KYSLGNTVWYDTNKDGIQGDDEKISGV--KVTLKDENGNIISTTTDENG---KIQPDN 768
Qy 297 LDRGHAVIDQ 308
Db 769 LNSGNTIVHFDK 780

```

## RESULT 9

```

O70022 STAEP PRELIMINARY; PRT; 1092 AA.
ID 070022 STAEP PRELIMINARY; PRT; 1092 AA.
AC 070022 STAEP PRELIMINARY; PRT; 1092 AA.
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DE 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Fibrinogen-binding protein precursor.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=HB;
RC MEDLINE=9826511; PubMed=9596732;
RA Nilsson M., Frydberg L., Flock J.I., Pel L., Lindberg M., Guss B.;
RT "A fibrinogen-binding protein of Staphylococcus epidermidis.";
RL Infect. Immun. 66:2666-2673(1998).
DR EMBL; Y17116; CAA7638.1; -; Genomic_DNA.
DR PIR; T30214; T30214.
DR HSSP; Q53653; 1N67.
DR SMR; O70022; 278-598.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR InterPro; IPR005877; GPro YSIRK.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF05738; Cna_B_2.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF04650; YSIRK_signal; 1.
DR TIGRFAMs; TIGR01167; LpxTG_anchor; 1.
DR TIGRFAMs; TIGR01168; YSIRK_signal; 1.
DR PROSITE; PSS0847; GRAM_POS_ANCHORING; 1.
KW Cell wall; signal.
FT STGNL 51 potential.
FT CHAIN 1 1092 fibrinogen-binding protein.
SQ SEQUENCE 1092 AA; 119292 MW; 6542BC39AAD8984 CRC64;

```

Query Match 7.0%; Score 124; DB 2; Length 1092;

Best Local Similarity 21.0%; Pred. No. 4.5;

Matches 78; Conservative 50; Mismatches 142; Indels 102; Gaps 16;

```

Qy 9 SDWTGVSTFVILETGLDNCNIYANGLMIG--VIINITPTDDEGNFVDIDVT-----60
Db 439 SMTNTIDTKNHYEQ-----TIYNPLRYSAKETNVAISGNGDEGSI-IDDSTIIIVYK 492
Qy 61 -----LNDNIKIVDYIDGSDIDGSDGMPYTGPNENYNT-----IPNS 97
Db 493 VGDNONLPDSNRRIYDYEVEDVT-NDQYAOQGNNDVNIANGNIDSPYIIKVISKVPNK 551
Qy 98 QASVSLKSENSQITQIRYVSCSTSLRTKSPAKY--TTTSG-----KVISI--144
Db 552 DDTVTTQOTVYMTTINYE-----TGEFRKASYNNTIAFSTSSQGGQDLPPEKTYKID 606
Qy 145 -----TONGSRRVAINADATNFTDELRTTKETRFENQSYTSHKSTNSL-Y 193

```

```

Db 607 YWEDVDKQIGQNTDNEKRLSNVLTLYTPDG---TSKSVRTDEBQKQPDGLKNGLT 663
Qy 194 VHTWTIPRSIKLQNMWEDY-----NNGWTAAQSCYRTGADGSEST 236
Db 664 KTFETP-----EGYPTLKHSQTNPALDSEGNVWVTINQDDMTIDSGFYQTP 713
Qy 237 RMLAAGSIFPPGNYDGLMNDIALSGMAHKSINVDGINQLSFTRIGGFSMVNNSG 296
Db 714 KYSLGNTVWYDTNKDGIQGDDEKISGV--KVTLKDENGNIISTTTDENG---KIQPDN 768
Qy 297 LDRGHAVIDQ 308
Db 769 LNSGNTIVHFDK 780

```

## RESULT 10

```

O931J9 GVIBR PRELIMINARY; PRT; 404 AA.
ID O931J9 GVIBR PRELIMINARY; PRT; 404 AA.
AC O931J9 GVIBR PRELIMINARY; PRT; 404 AA.
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DE 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein.
OS Vibrio pomeroyensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=161725;
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=CH-291;
RC PubMed=14680703; DOI=10.1016/S0378-1097(03)00843-7;
RA Jores J., Appel B., Lewin A.;
RT "Cloning and molecular characterization of a unique hemolysin gene of
RT Vibrio pomeroyensis sp. nov.: development of a DNA probe for the
RT detection of the hemolysin gene and its use in identification of
RT related Vibrio spp. from the Baltic Sea.";
RL FEMS Microbiol. Lett. 229:223-229(2003).
DR EMBL; AJ314791; CAC40977.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 404 AA; 45760 MW; 0EB685219A4AC331 CRC64;

```

Query Match 6.9%; Score 122.5; DB 2; Length 404;

Best Local Similarity 21.1%; Pred. No. 1.7; 127; Indels 131; Gaps 20;

Matches 85; Conservative 59; Mismatches 142; Indels 131; Gaps 20;

```

Qy 14 VSTFVILETGLDNCNIYANGLMIG--VIINITPTDDEGNFVDIDVT-----60
Db 24 IKDPSISL--SGSDRGQIFSGNNGNIBFVN-----VDLDEGEIEIWKILRKQDK 72
Qy 74 SDIDGSDGMPY--TGNPNEVNTI-----PNSQSYSLKSR-----N 107
Db 73 KDIIQ-ELGWLRYRPLQSDSEYYSIHGKYALPNAFIRTTQGNPFAQSEILSKADQSIQYN 131
Qy 108 SQTQIKRY-----VSCSNTSR-----RTKFSKAYTTTSGV--ISTONGSINS 151
Db 132 SKLQSIENYETPLPTVTHSQTLVLSQDVQVDCAEVTLKSGVLYDSCEVSEPTTV 191
Qy 152 SRVAINADATNFTDELRTTKETRFEN--QSYTSHKS--STNSL--YVHTWTIPRSIKL 205
Db 192 SSTTPVIEPENGSSSDLDADIDIEVNNQHSSSLKALAPALABEFIAIYIDSDDE 251
Qy 206 QNRWEDYNGWTAAQSCYRTGADGSESTRLAAGSIFPPGNYDGLMND--IALS 262
Db 252 PLIYDYEYKVPAGEKTAAY-----NVDGINQLSFTRIGGFSNV-----WIDNETYHPLP 283
Qy 263 GMAHKS--NVDGINQLSFTRIGGFSNV-----WIDNETYHPLP 283
Db 284 NUNNFKTLTGRLDELAKLSVSSKINTLT-----RGFTIILNVVTTIPNMQSYDTDTW 337
Qy 297 LDRGHAVIDQ 308
Db 338 VVDG--YVNTTYGGCYFPPERNPLQANSNTCKLKLSEIL 376

```

RESULT 11  
Q97VMO\_SULSO PRELIMINARY; PRT; 990 AA.  
AC Q97VMO;  
DT 01-OCT-2001 (TREMBlrel. 18, Created)  
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)  
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
DE Hypothetical protein.  
GN OrderedLocustNames=SSO2602;  
OS Sulfolobus solfataricus.  
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
OC Sulfolobus.  
NCBI\_TaxID=2287;  
RN NUCLEOTIDE SEQUENCE.  
RP STRAIN=ATCC 35092 / DSM 1617 / P2;  
RX MEDLINE=21332296; PubMed=11427726; DOI=10.1073/pnas.141222098;  
RA Shie Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,  
RA Aweez M.J., Chan-Weiner C.C.-Y., Clausen I.G., Curtis B.A.,  
RA De Moor A., Eiraso G., Fletcher C., Gordon P.M.K.,  
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,  
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,  
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,  
RA Garrett R.A., Regan M.A., Senses C.W., Van der Oost J.,  
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.",  
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).  
DR EMBL; AE006856; AAK42724.1; -; Genomic\_DNA.  
DR PIR; E90433; E90433.  
KM Complete proteome; Hypothetical protein.  
SQ SEQUENCE 990 AA; 111429 MW; DBE2517A7FBA596A CRC64;

Query Match 6.9%; Score 122.5; DB 2; Length 990;  
Best Local Similarity 20.8%; Pred. No. 5.1;  
Matches 64; Conservative 45; Mismatches 99; Indels 99; Gaps 13;  
QY 36 NMIVIIITPTDEGNFVDID-----VLTNDIKIVYIDGSDIDGSDGW 82  
DB NLVVISLQKPTSN--NVTIYNGPVSLKHYVLYQYINLISNIPKALINLTIVTFSGW 785  
QY 83 FYTGNP---NEYTIIPNSQSYSL-----KSENSQITQIKRVVSCSNTSRRLRTK 128  
DB YNAVTPISFIMTYIISNTRYIILSLPFPVNRSLTVAVTLTKEL----- 834  
QY 129 SPSAKVTTTSGKAVSITONSINSKRVVNAIDATNFTDDEKRTKTRFEQOSTSHKSS 188  
DB VTNNEPILATIANRINTNTSEIWPV-----CQTLIPKYVVISNBERIFYNTS 882  
QY 189 TNSLYVHWTLPRSLKQNMWRMEDYNNGMWTAQSCYVTGADGSESTRMLAAGS----- 243  
DB S-----YLNIITQPTISNVK-----PIIEYVTTIDGNS--WIPRGSVITLT 922  
QY 244 -----IFPPGNDYGLMDNDIALSGMAHKSYNVDGTGINSFTRIIIGKFSWYVNIIGDR 299  
DB QSVPIVEGKNEG-----SYNVSNGV-AITVNOPTTERF--VKQINSGFV 964  
QY 300 GHAVIII 306  
DB 965 GSVIII 971

RESULT 12  
Q7V8S5\_PROMM PRELIMINARY; PRT; 1765 AA.  
AC Q7V8S5;  
DT 01-OCT-2003 (TREMBlrel. 25, Created)  
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)  
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
DE Hemolysin-type calcium-binding region: RTX N-terminal domain.  
GN OrderedLocustNames=PMT0256;  
OS Prochlorococcus marinus (Strain MIT 9313).  
OC Bacteria; Cyanobacteria; Prochlorales; Prochlorococcaceae;

OC Prochlorococcus.  
OX NCBI\_TaxID=74547;  
RN NUCLEOTIDE SEQUENCE.  
RP MEDLINE=22825698; PubMed=12917642; DOI=10.1038/nature01947;  
RA Rocap G., Larimer F.W., Lamerdin J.E., Malafati S., Chain P.,  
RA Ahlgren N.A., Arellano A., Coleman M., Hauser L., Heas W.R.,  
RA Johnson Z.I., Land M.L., Lindell D., Post A.F., Regala W., Shah M.,  
RA Shaw S.L., Steglich C., Sullivan M.B., Ting C.S., Tolonen A.,  
RA Webb E.A., Zinner E.R., Chisholm S.W.;  
RT "Genome divergence in two Prochlorococcus ecotypes reflects oceanic  
niche differentiation.";  
RL Nature 424:1042-1047(2003).  
DR EMBL; BX572095; CAE20431.1; -; Genomic DNA.  
DR GO; GO:0005576; C:extracellular region; IEA.  
DR GO; GO:0005509; F:calcium ion binding; IEA.  
DR GO; GO:0005529; F:sugar binding; IEA.  
DR GO; GO:0009405; P:pathogenesis; IEA.  
DR InterPro; IPR001343; Hemolysn\_Ca\_bind.  
DR InterPro; IPR001304; Lectin\_C.  
DR InterPro; IPR003995; RtxA.  
DR Pfam; PF00353; Hemolysincbind; 5.  
DR Pfam; PF00059; Lectin\_C\_1.  
DR PRINTS; PR00313; CMBNDNGRPT.  
DR PRINTS; PR01488; RTXTOXINA.  
DR PROSITE; PS00330; HEMOLYSIN\_CALCITUM; 1.  
KM Complete proteome.  
SQ SEQUENCE 1765 AA; 187603 MW; 817CE4F5007580CD CRC64;

Query Match 6.8%; Score 121; DB 2; Length 1765;  
Best Local Similarity 20.1%; Pred. No. 13;  
Matches 75; Conservative 60; Mismatches 146; Indels 92; Gaps 14;  
QY 4 REDSMRWGVSTFNVILEHGLDNCNIYAN-GLNMIGVLIINPTDEGNFVDIDVTYN 62  
DB 705 RHISYGATGADGNSITSGSDSVNLQINAGSALGNDLSITSDGNDITDITAIIG 764  
QY 63 DNIIKIVYIDGSDIDGSDGFFYTGNP-----EYNTIIPNSQSYSLKSENSQITQ--IKR 115  
DB 765 ENS-----FYKSSSSRSRSPDSTGNSSDYRIISRSRSGSYNDYTN 807  
QY 116 YVSCSNTSRRLTKSPSAKVTTTSGKAVSIT-----QNSINSRVVNAIDATN 163  
DB 808 YSNSYERSDASSSFNDSSTNSRSIGQSTYKNSYSGYSGWYNSSPYERSHLNLSY--YD 866  
QY 164 FTIDELRTTETRENSY---TSHKSTNSLYVHWTLIPRSLKQNMWR--EDYNNGW 218  
DB 867 YSNSSSSSSHNSNSYSRYNDNTRYSNSSDYLNSSHQSSYDNNRFQSSYDYDHS 926  
QY 219 WAQSCYKGTADG-----GSESTR--WLAAGSIFPPGN-----YDGLWLDNDIAL-- 261  
DB 927 RHISYGATGADGNSITSGSDSVNLQINAGSALGNDLSITSDG--NDDITIDVN 983  
QY 262 -----SGMAHKSYNVDGTGINSFTRIIIGKFSWYVNIISGLRGHVVITIDQYN 311  
DB 984 AFGENSFYKSSSRSSRSPDSTGNSSDYSISYRSSS-----YSN 1024  
QY 312 KYRIIFHAGYENS 324  
DB 1025 DYTNSYSNSYDHS 1037

RESULT 13  
O52984\_ECOLI PRELIMINARY; PRT; 814 AA.  
AC O52984;  
DT 01-JUN-1998 (TREMBlrel. 06, Created)  
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Hypothetical protein precursor.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.

OX NCBI\_TaxID=562;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE  
 RX MEDLINE=99412288; PubMed=10482530;  
 RA Dutky T.G., Steadher L.H., Manning P.A.;  
 RT "55 Plus Bioynthesis Genes from Enterotoxigenic Escherichia coli  
 115:H40";  
 RL J. Bacteriol. 181:5847-5851(1999).  
 DR EMBL: AJ224079; CA11822.1; -; Genomic DNA.  
 DR GO: GO:0016020; C:membrane; IEA.  
 DR GO: GO:0005215; F:transporter activity; IEA.  
 DR GO: GO:0006810; P:transport; IEA.  
 KM Hypothetical protein; Signal.  
 FT SIGNAL 1 21 Potential.  
 FT CHAIN 1 1 hypothetical protein.  
 FT CHAIN 22 814 hypothetical protein.  
 SQ SEQUENCE 814 AA; 90322 MW; C2108P2461BD892 CRC64;  
  
 Query Match 6.7%; Score 119; DB 2; Length 814;  
 Best Local Similarity 20.0%; Pred. No. 7.2;  
 Matches 94; Conservative 62; Mismatches 141; Indels 172; Gaps 22;  
  
 QY 1 FTLEDSMSDWTGVSTFNV-----ILFTGLDNCNIIA-----NG 34  
 DB FGTAEELVDRSYGKTFSSYSPSTGIVEKDNILVLAITQAGYGEINLANLPYQYNA 326  
 QY 35 LNM-----GIITITPTDE-----GPFVIDDVTLDNIIKVD----- 69  
 DB 327 LVQVSSSGIVSSQNLINNTGSFNSDFSWHLFVGNSSGSDNEFVKTEVEISGVLP 386  
 QY 70 -----YIDGSDIDSGDFYTGPNBEVNTI-----PNSQSYSLKSENQITQ 112  
 DB 387 VNTLTALYVGAKVD-----KTIITYTGMPQKEPISVLRKGGGQGFPHYE 433  
 QY 113 IKRVSCSNTSRLT-----KSFSAKVTTSKGVISITON-SINSSRVVINAIDAT 162  
 DB 434 MKSYLEBLSLMKKTSIGKMNGLKSLDNTTLISAGYNFVMSVSNVGVYSSWRPD 493  
 QY 163 NF--TDELRTTETREPNOSYTSHK-----SSTNSLYHTWTIIPPSLQNRME 211  
 DB 494 YFYANTDHLMESEFRKKNYSNKL YANMYNFPGNLSLYMTY-----KELRGN 545  
 QY 212 DY-----NNGMTWAQSCYKTDGSGEST-----RWLAAGSIFPPGNY 250  
 DB 546 DYISVSLGMITSLGNRPF--NSGFYKNGADITMSITVDYAKSLDMWSHVSV----- 596  
 QY 251 DGLWLDND--IALSGNAHKS-----YNDVTGINQLSF--TRIIGK-----GP 288  
 DB 597 -GRFSDSYNSATYSLSHNSNEVRGAGYYATDNGSQLTLPLADSTQIINSNGIYFTSS 655  
 QY 289 SWTYN--ISGLDRGHAVIIIDYGNKRIILFHAGYENSDEPLSSIVY 334  
 DB 656 SWKDNAPFIRGDKAKYDISVNMNTDNTTR-----YDSDPTNIISVEVY 698  
  
 RESULT 14  
 O832P7\_ENTPA PRELIMINARY; PRT; 893 AA.  
 ID O832P7  
 AC O832P7  
 DT 01-JUN-2003 (TREMBlrel. 24, Created)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE Conserved domain protein.  
 GN OrderedLocustNames=SP2174;  
 OS Enterococcus faecalis (Streptococcus faecalis).  
 OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.  
 OX NCBI\_TaxID=1351;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=V583 / ATCC 700802;  
 RX MEDLINE=22550857; PubMed=12663927; DOI=10.1126/science.1080613;  
 RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seshadri R.,  
 RA Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,

RA Tettelin H., Dodson R.J., Umayam L.A., Brinkac L.M., Beanan M.J.,  
 RA Dougherty S.C., Deboy R.T., Durkin S.A., Kolony J.F., Madupu R.,  
 RA Nelson W.C., Vamathevan J.T., Tran B., Upton J., Hansen T., Shetty J.,  
 RA Knouri H.M., Uetereback T.R., Radue D., Ketchum K.A., Dougherty B.A.,  
 RA Fraser C.M.;  
 RT "Role of mobile DNA in the evolution of vancomycin-resistant  
 RT Enterococcus faecalis";  
 RL Science 299:2071-2074(2003).  
 DR EMBL: AE016953; MA081906.1; -; Genomic DNA.  
 DR TIGR: EF2174;  
 DR GO: GO:0003796; F:lysozyme activity; IEA.  
 DR GO: GO:0016998; P:cell wall catabolism; IEA.  
 DR GO: GO:0009253; P:peptidoglycan catabolism; IEA.  
 DR InterPro: IPR010253; Glyco hydro 25.  
 DR InterPro: IPR010916; TONB\_Box\_N..  
 DR Pfam: PF01183; Glyco\_hydro\_25; 1.  
 DR ProDom: PD004620; Glyco\_hydro\_25; 1.  
 DR PROSITE: PS00430; TONB\_DEPENDENT\_REC\_1; UNKNOWN\_1.  
 KM Complete proteome.  
 SQ SEQUENCE 893 AA; 99579 MW; BCE7D9764AE80DB CRC64;  
  
 Query Match 6.7%; Score 119; DB 2; Length 893;  
 Best Local Similarity 19.4%; Pred. No. 8;  
 Matches 74; Conservative 58; Mismatches 119; Indels 130; Gaps 20;  
  
 QY 21 LFTGLDN-----CNIVANGLMIGVIINITPTDEGNFVID-----DVTLDNIIKVD 69  
 DB 138 MKNGLSRVARATVAVNYADPNLPG-----KNITDVSNNGDISVAEQIKS 186  
 QY 70 Y-IDGSDIDSGDFYTGPNBEVNTI-PNSQSYSLKSENQITQIKRVYSCSNTSRLTK 128  
 DB 187 YGVTVGVSKLTETWYV-NPYAAGQIRMAKAALKYS-----AHYSYVSA-A-TAQDEAR 240  
 QY 129 SPSAKVTTSKGVISITONSINSSRVVINAIDATNFTDELRTTETREPNOSYTSKSS 188  
 DB 241 YF-AQAANSNG-----LDKNTIMFNDAPETLTNNRMAHNSVAFNODLKALGYK 290  
 QY 189 TNSLYHTWTI-----PSLSLQNRWEDYNNGMWMAQSCY 225  
 DB 291 NDALYVGKMWLTNGYIDTSAFGDRVVAQYPTTPDS--SMQNDHGAQMSQMTF 346  
 QY 226 KTGADGSGESTRWLAAGSIFPPG--NYDGLMDNDIALSGMAH-----KSY----- 269  
 DB 347 -----PGLANYEGRPPIISMYSFAMGNSSGSDLSKYITNP 385  
 QY 270 -----NVDGTINQLSF-----TRIIGFSGVNTNISGLDRGHAVIITD--YG 310  
 DB 386 GRVIMKNDPTFYQDVAFRTPGWRKKNITLVIRGIR--YSSAGIPR-----LVTDQGYLTA 439  
 QY 311 NKYRIILFHAGYENSDEPLSS 331  
 DB 440 NKDYVL--AAQSNIDLYFTTN. 458  
  
 RESULT 15  
 Q4L555\_STRAJ PRELIMINARY; PRT; 1361 AA.  
 ID Q4L555  
 AC Q4L555  
 DT 13-SEP-2005 (TREMBlrel. 31, Created)  
 DT 13-SEP-2005 (TREMBlrel. 31, Last sequence update)  
 DT 13-SEP-2005 (TREMBlrel. 31, Last annotation update)  
 DE Autolysin.  
 GN Name=atl; ORFNames=SH1911;  
 OS Staphylococcus haemolyticus (strain JSC1435).  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 OX NCBI\_TaxID=279808;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=JSC1435;  
 RA Takeuchi Y., Watanabe S., Baba T., Yuzawa H., Ito T., Cui L.,  
 RA Morimoto Y., Kuroda M., Takahashi M., Anai A., Baba S., Fukui S.,  
 RA Lee J.C., Hiratake K.;  
 RT "Whole genome sequencing of Staphylococcus haemolyticus uncovers



GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: January 30, 2006, 09:40:03 ; Search time 128.372 Seconds  
(without alignments)  
5726.190 Million cell updates/sec

Title: US-09-889-874a-23

Perfect score: 8879  
Sequence: 1 VYIKFLKLFRRITMSDNNF.....PKIIIGRETKVTKPTFR 1673

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues  
Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : A\_Geneseq\_21:\*

- 1: geneseqp1980s:\*
- 2: geneseqp1980s:\*
- 3: geneseqp2000s:\*
- 4: geneseqp2001s:\*
- 5: geneseqp2002s:\*
- 6: geneseqp2003as:\*
- 7: geneseqp2003bs:\*
- 8: geneseqp2004s:\*
- 9: geneseqp2005s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8879	100.0	1673	3	AAV95707
2	1638	18.4	1584	2	AAV95707
3	1601	18.0	1410	8	ADQ26337
4	1583	17.8	1787	6	ABM67433
5	1569	17.7	1590	6	ABM67283
6	932.5	10.5	982	7	ADG73113
7	932.5	10.5	982	7	ADL12168
8	517.5	5.8	1317	8	ADS22210
9	432	4.9	2234	5	ABG31849
10	386	4.3	2234	6	ABU18641
11	365	4.1	1400	7	ADC01365
12	361	4.1	1510	7	ADG30698
13	353.5	4.0	843	6	ABU48290
14	350	3.9	1515	6	ABU19676
15	342	3.9	1397	6	ABU15135
16	341.5	3.8	1404	7	ADC00960
17	341	3.8	1394	7	ADC01426
18	340.5	3.8	1411	6	ABU15136
19	339	3.8	1426	6	AAH51983
20	339	3.8	1426	6	ABU14693
21	331.5	3.7	1377	4	AAU34791
22	331.5	3.7	1377	4	AAU34791
23	331.5	3.7	1377	6	ABU14807
24	328	3.7	1572	6	ABU41491

25	324.5	3.7	1429	6	ABM69806	Abm69806
26	322	3.6	1385	6	ABU40318	Abu40318
27	317.5	3.6	1504	6	ABM66973	Abm66973
28	317.5	3.6	1565	6	ABU16634	Abu16634
29	317	3.6	998	7	ABO79678	Abu79678
30	314	3.5	1395	6	ABU22662	Abu22662
31	313	3.5	1627	6	ADA35317	Ada35317
32	307.5	3.5	932	6	ABU16693	Abu16693
33	306.5	3.5	1530	6	ABU40251	Abu40251
34	306	3.4	1566	7	ADF05044	Adf05044
35	298.5	3.4	1531	6	ABU21384	Abu21384
36	298	3.4	885	4	AU52772	Au52772
37	298	3.4	885	6	ABM49291	Abm49291
38	298	3.4	920	6	ABM65127	Abm65127
39	297	3.3	1512	6	ABU50462	Abu50462
40	294	3.3	1439	7	ADF07275	Adf07275
41	292	3.3	1626	7	ABO75059	Abu75059
42	291.5	3.3	1253	6	ABU40990	Abu40990
43	291	3.3	2346	5	ADL16955	Adl16955
44	290.5	3.3	2802	5	ADL16953	Adl16953
45	289.5	3.3	1364	6	ABU47385	Abu47385

## ALIGNMENTS

RESULT 1  
ID AAV95707 standard; protein; 1673 AA.  
XX AAV95707;  
AC AAV95707;  
DT 25-OCT-2000 (first entry)  
XX Cosmid cHRIM5 encoded protein P14-2F.  
DE Cosmid cHRIM5; nematocyst; nematode; biological control agent;  
XX Cosmid cHRIM5; nematocyst; nematode; biological control agent;  
KW transgenic plant; helminthiasis; P14-2F.  
XX Xenorhabdus bovienii.  
OS WO200042855-A1.  
PN 27-JUL-2000.  
PD 24-JAN-2000; 2000WO-GB000219.  
PE 22-JAN-1999; 99GB-00001499.  
PR (HORT-) HORTICULTURE RES INT.  
PA Morgan JAW, Jarrett P, Ellis D, Ousley MA;  
XX WPI; 2000-499157/44.  
XX N-PSDB; AAA50029.  
XX Novel composition used to control parasitic nematodes, especially in a plants such as maize, cotton, soya, and rice, comprises a bacterium which is a symbiont of an entomopathogenic nematode.  
PS Example 6; Page 42-43; 74pp; English.  
XX The present sequence is that of protein P14-2F encoded by an open reading frame identified in cosmid cHRIM5 (see AAA50029). cHRIM5 was obtained by ligating Xenorhabdus bovienii strain I73 (NCIMB 40986) Sau3A-digested DNA fragments into the BamHI site of the Stratagene cosmid vector Supercoiled, packaging into Escherichia coli XL Blue 1, and screening for nematocidal activity against Caenorhabditis elegans. Analysis of the DNA indicated a number of open reading frames for which the corresponding protein sequences were determined (see AAV95685-Y95735). Nematodes can be controlled through the use of bacteria associated symbiotically with an entomopathogenic nematode. Such bacteria include Xenorhabdus and Photobacterium spp. such as X. bovienii strain I73. The symbiont bacteria,

CC an engineered bacterium, or a nematocidal protein obtained from such  
 CC bacteria, particularly p13-1f (see AAY95706) or p14-2f can be used to  
 CC control helminths in a human or domesticated animal or for the  
 CC control of plant pathogen nematodes. Also claimed are vectors for  
 CC expressing nematocidal proteins in host cells, and transgenic plants  
 CC

XX Sequence 1673 AA;

Query Match 100.0%; Score 8879; DB 3; Length 1673;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1673; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYIKFLKFRITWSDNNEFFTOANNFTSAVSGVDPRGLYNIQITLHIVGNGLGPT 60  
 DB 1 VYIKFLKFRITWSDNNEFFTOANNFTSAVSGVDPRGLYNIQITLHIVGNGLGPT 60  
 QY 61 LPLTLVSPLNKTDIGFGIGFNLGYDRKNSILSLSTGENYKVIETDKTVKLOQKLD 120  
 DB 61 LPLTLVSPLNKTDIGFGIGFNLGYDRKNSILSLSTGENYKVIETDKTVKLOQKLD 120  
 QY 121 NLREKOLKENCYRIHKSGLIEVLTFGNNAFLKPKKLLNPAHAITYIDMFEATOP 180  
 DB 121 NLREKOLKENCYRIHKSGLIEVLTFGNNAFLKPKKLLNPAHAITYIDMFEATOP 180  
 QY 181 RLNRIDYDLGDHDIPLNLBYOGLIKTILFLPGQKSGYRTELRLNROLNSIHNFSLGN 240  
 DB 181 RLNRIDYDLGDHDIPLNLBYOGLIKTILFLPGQKSGYRTELRLNROLNSIHNFSLGN 240  
 QY 241 ENPLTWSEGYTPIGKNGILGOWITSMAPGGLKXTVYNNNOGHHPQSANLEVLVYT 300  
 DB 241 ENPLTWSEGYTPIGKNGILGOWITSMAPGGLKXTVYNNNOGHHPQSANLEVLVYT 300  
 QY 301 LMKVPGAGOPALIAEVSYSHTNVYGGSGNGIMNNKLDNLGLMTEYVYGSTESRYKDK 360  
 DB 301 LMKVPGAGOPALIAEVSYSHTNVYGGSGNGIMNNKLDNLGLMTEYVYGSTESRYKDK 360  
 QY 361 EGHQOIVARIERTVNNYHLITSECKQONGYIQTETAYYAIIGHNPDSPQSFOLPKYTE 420  
 DB 361 EGHQOIVARIERTVNNYHLITSECKQONGYIQTETAYYAIIGHNPDSPQSFOLPKYTE 420  
 QY 421 TWRADNSYRSEIETETTFDESIGNPLTKYIKDKTKQKIIISPTHWEEYYPAGEVNCPEP 480  
 DB 421 TWRADNSYRSEIETETTFDESIGNPLTKYIKDKTKQKIIISPTHWEEYYPAGEVNCPEP 480  
 QY 481 YGFRFVVKKIIOTPRDSEFKDPEKFIQRYSLIGSOSHVLTKIEHRYSATQLNSTLF 540  
 DB 481 YGFRFVVKKIIOTPRDSEFKDPEKFIQRYSLIGSOSHVLTKIEHRYSATQLNSTLF 540  
 QY 541 QYNTDKSELGRLLKQTECTKSGNGKTVSVHKFTYTKODDTLQOSSHITTHDNFTIHSQ 600  
 DB 541 QYNTDKSELGRLLKQTECTKSGNGKTVSVHKFTYTKODDTLQOSSHITTHDNFTIHSQ 600  
 QY 601 VRSRYTGLFSDPTDXTDVIYQMSYDKLGRLLTRTLNSGTPYANTLTYDELNLODDNR 660  
 DB 601 VRSRYTGLFSDPTDXTDVIYQMSYDKLGRLLTRTLNSGTPYANTLTYDELNLODDNR 660  
 QY 661 PVRSTYGLRFDPTDXTDVIYQMSYDKLGRLLTRTLNSGTPYANTLTYDELNLODDNR 660  
 DB 661 PVRSTYGLRFDPTDXTDVIYQMSYDKLGRLLTRTLNSGTPYANTLTYDELNLODDNR 660  
 QY 721 LTNQOQTDPDVKVHLSMSKSYDNMGOIANTHWSYVSEKITYDPTTLTATKOLQSNNSNV 780  
 DB 721 LTNQOQTDPDVKVHLSMSKSYDNMGOIANTHWSYVSEKITYDPTTLTATKOLQSNNSNV 780  
 QY 781 QTEKEVTTYPSQOPIQITLFDAGHLOSCHTLTFDGMVRKETAIGQCTIYYDNYN 840  
 DB 781 QTEKEVTTYPSQOPIQITLFDAGHLOSCHTLTFDGMVRKETAIGQCTIYYDNYN 840  
 QY 841 RVIOQTTLPGTIVNRKVAPEFSTDTLITDRVNGISLGGQTFPGLSRLTQSDGGGVMAVT 900  
 DB 841 RVIOQTTLPGTIVNRKVAPEFSTDTLITDRVNGISLGGQTFPGLSRLTQSDGGGVMAVT 900  
 QY 901 YSAGNDQCPSTVITPDGFIHYQOPELDDAVLQVANSNEITQOFSYNPVTGALLKAVAE 960

DB 901 YSAGNDQCPSTVITPDGFIHYQOPELDDAVLQVANSNEITQOFSYNPVTGALLKAVAE 960  
 QY 961 OSITPIYPSGRKMEINMDMKMSYLMTLRGLHNGYTDLTGTIOKISRDTGRVTOIKD 1020  
 DB 961 OSITPIYPSGRKMEINMDMKMSYLMTLRGLHNGYTDLTGTIOKISRDTGRVTOIKD 1020  
 QY 1021 SSIKTTLANDDLNRHIGSOVTDLATGMLTTTVEPDLNREIGRKLCDSSGHTLDIQSW 1080  
 DB 1021 SSIKTTLANDDLNRHIGSOVTDLATGMLTTTVEPDLNREIGRKLCDSSGHTLDIQSW 1080  
 QY 1081 LKTQOLANRIVKLVGLQRTBOYSYSRRLNRYKCDGACCPYDKGHSIVTQNFYDIY 1140  
 DB 1081 LKTQOLANRIVKLVGLQRTBOYSYSRRLNRYKCDGACCPYDKGHSIVTQNFYDIY 1140  
 QY 1141 GNITACHTTPADGTEDHATKPFANPDPCCOLTEVHTHPMPNIRLKYDKAGRVINITD 1200  
 DB 1141 GNITACHTTPADGTEDHATKPFANPDPCCOLTEVHTHPMPNIRLKYDKAGRVINITD 1200  
 QY 1201 NHGNTENFTYDTLGRLONGGSVYGYDPLNRLVSOKTDLDCSLYYRETMVNEVRNGEM 1260  
 DB 1201 NHGNTENFTYDTLGRLONGGSVYGYDPLNRLVSOKTDLDCSLYYRETMVNEVRNGEM 1260  
 QY 1261 IRLLRGETTIAOQRAKVLITGTDQOQSVILTSQKNSQBAVSAVGRKSTANDASIL 1320  
 DB 1261 IRLLRGETTIAOQRAKVLITGTDQOQSVILTSQKNSQBAVSAVGRKSTANDASIL 1320  
 QY 1321 GYNGERADPVSGVTHLNGRYSYDPTLMRFPTPDSLSPFAGAGINPYCYLADPINSRP 1380  
 DB 1321 GYNGERADPVSGVTHLNGRYSYDPTLMRFPTPDSLSPFAGAGINPYCYLADPINSRP 1380  
 QY 1381 SGHLSQWAMTGIQWINGLGLLTATGMAIAAGGIAAIAASTTALAFALSVTSIDT 1440  
 DB 1381 SGHLSQWAMTGIQWINGLGLLTATGMAIAAGGIAAIAASTTALAFALSVTSIDT 1440  
 QY 1441 SYVSGALEDSAPRASSTLGVNMGMAAGLAESAIKGTGLATHLGAFADGENALLKST 1500  
 DB 1441 SYVSGALEDSAPRASSTLGVNMGMAAGLAESAIKGTGLATHLGAFADGENALLKST 1500  
 QY 1501 SESSRIKMGVTRSLDREIVANBEGQVTKDSRGTYTDFPMKGBQALIVHGDQDFLYHTE 1560  
 DB 1501 SESSRIKMGVTRSLDREIVANBEGQVTKDSRGTYTDFPMKGBQALIVHGDQDFLYHTE 1560  
 QY 1561 GNKNGKGPYTRHPEGLVDYLDKNNIVDLTGGDKPVLHLSYCKSSGAADKMAKTYNR 1620  
 DB 1561 GNKNGKGPYTRHPEGLVDYLDKNNIVDLTGGDKPVLHLSYCKSSGAADKMAKTYNR 1620  
 QY 1621 PVIAYSNKPITISQGLARIERKDPFLKSTYHSYDPKRIILGRTEKTVKPTFRP 1673  
 DB 1621 PVIAYSNKPITISQGLARIERKDPFLKSTYHSYDPKRIILGRTEKTVKPTFRP 1673

RESULT 2  
 AAY33727  
 ID AAY33727 standard; protein; 1584 AA.  
 XX  
 XX AAY33727;  
 DT 09-NOV-1999 (first entry)  
 XX  
 XX Photobhabdus luminecens 176 kD insecticidal toxin.  
 DE Photobhabdus luminecens 176 kD insecticidal toxin.  
 KW Symbolic bacterium; nematode; insect; larva; toxin; insecticide.  
 OS Photobhabdus luminecens.  
 XX  
 XX W09942589-A2.  
 XX  
 XX 26-AUG-1999.  
 PD 26-AUG-1999.  
 XX  
 XX 18-FEB-1999; 99MO-EP001015.  
 PF 18-FEB-1999; 99MO-EP001015.  
 XX  
 XX 20-FEB-1998; 98US-00027080.  
 PR



PR 20-JAN-1999: 99US-0116439P.  
 XX (NOVS ) NOVARTIS AG.  
 PA (NOVS ) NOVARTIS-ERFINDUNGEN VERN GES MBH.  
 XX Kramer VC, Morgan MK, Anderson AR, Hart HP, Warren GM, Dunn MM;  
 PI Chen JS;  
 XX WPI: 1999-527479/44.  
 DR N-PSDB: AA206826.  
 PT New nucleic acid from Photobacterium luminescens encoding insecticidal  
 XX toxin, used for making resistant transgenic plants.  
 XX  
 PS Claim 25; Page 93-98; 148pp; English.  
 CC This sequence represents an approximately 176 kD insecticidal toxin from  
 CC Photobacterium luminescens. It is one of five insecticidal toxins (AA23723  
 CC -Y33727) encoded by open reading frames (orfs) arranged in an operon-like  
 CC structure in a 9.7kb fragment of P. luminescens DNA (AA206826). This  
 CC sequence is encoded by orfs. P. luminescens is a member of the  
 CC Enterobacteriaceae family and is a symbiotic bacterium of nematodes of  
 CC the genus Heterorhabdus. The nematodes colonise insect larvae, kill  
 CC them, and their offspring feed on the dead larvae. However, the  
 CC insecticidal agents are produced by P. luminescens rather than the  
 CC nematode. The toxins have activity against lepidopteran insects such as  
 CC cabbage looper (Trichoplusia ni), European corn borer (Ostrinia nubilalis)  
 CC and Fall Armyworm (Spodoptera frugiperda) and also against Coleopteran  
 CC insects (e.g., Colorado potato beetle, Leptinotarsa decemlineata). In  
 CC addition the toxins are active against strains resistant to known  
 CC insecticides. The DNA sequence can be used to generate transgenic plants  
 CC of various species that are resistant to economically important insect  
 CC pests and also for recombinant production of the toxins for use as  
 CC insecticides  
 XX  
 SQ Sequence 1584 AA;  
 Query Match 18.4%; Score 1638; DB 2; Length 1584;  
 Best Local Similarity 31.0%; Pred. No. 4.6e-105;  
 Matches 493; Conservative 233; Mismatches 682; Indels 182; Gaps 54;  
 QY 19 EFTQANNTSAVSGVDPRTGLNIIQITLGHVGN--GNLGPPLPLTSLVSPPLAKTDIG 76  
 Db 5 DIYSNAPNFGSYINTGVDPRTGYSANINIIILAPNVNSEQ--LSLSFSLPTTLANG 62  
 QY 77 FGIGFNGSLSVYDRKNSLSLSTGENYK---VETDKTVLQCKLDMNLFKDKLKNKY 133  
 Db 63 FGIGMRFSLLTLLDKITLTFRRANGEQFCKPLPNNNDLSFKDKLADLRYK-LDSNTF 121  
 QY 134 RIHKSGDIEVLGPNNAFDLAKPKGLNPAAGAIYIDNFFAOTRRLRIYDDLDGHD 193  
 Db 122 YVNNKGIIILIKRIGSS--DIAKTVALPEPDGAFLIYNSR----- 162  
 QY 194 IPLNLVEYQGLIKTILTL-PPQKEGRTFLRL-NQLNSIHFLSGNENPLTWSFGYT 251  
 Db 163 FALSEIKYRTGKTYLKLANSNN--CTSEVYEDDNNISAKAFDVRANDYILITVTPYD 219  
 QY 252 PIQK-----NGILGQWITSMTAPGLKETVANSNNQGHHPQSANLPVLPY 298  
 Db 220 ASGPIDISARFKMTYQTLKGFV-PVISTFRTPTGVELVSYKEN--GH--KVTDTETIPY 273  
 QY 299 VTLMKOVPGAGPAIQAEVYS--HNVVGGSGNCL--WNMLDMLYGLMTEYNGSTSR 356  
 Db 274 AALATLTPGNGGPAVSVSYSSVHNPLVYSGRTSDSSQDNLYLTGKTYSSLE-- 330  
 QY 357 YKDEGHDIQVIRERTNNYHLTSECKQNGYIQTETAYVAILIGHNPSQSPQLP 415  
 Db 331 -RVLDDGSVSVIERVNNKFLMTKEAKTDNKRITTEITTYNEDLSGSPQENLQPS 369  
 QY 416 KYTEETRSADNSYRSEITETTPDESQNPITKVIKDKTKKTIISPTSHWEYPPAGEVDN 475  
 Db 390 RVLTRYTDIQTNTSRRETVNIKSDMGNTLL-ITETSGIOL-----EYVYVPGEGNS 442

QY 476 CPPEPVGTRFVKKIITQPYDSEKDDPEKEIOYR-----SLIGSQ-----SHVTLKIEE 526  
 Db 443 CPADPLGFSRFLSVYOKSGPDAAQSVANKVITHYTKFPTFGAIVKEVSKVSEITDN 502  
 QY 527 RHYSAQLNLSLTFQYNTDKSELGRLLKQTECKGENGKTVSVVHKFTYTKODDTLQOSH 586  
 Db 503 KIARTFSYVNSP-----TSKSH--GSLAKITSVMNNQ-----QVWTFKYESESEMTNA 551  
 QY 587 SITTHDNFTIHRGQVRRYRGLPSDITDITQMSYDGLGRLLTTLTASGTPYANTLT 646  
 Db 552 TVTGFDAHMSKNVNTSYTHRQLRKVDVNVHTDQSYDLGRITGQIIPGTAREIKRN 611  
 QY 647 YDEILNLQDNPPFITTTTVDVNGQLRPEFGAGHVSQCLKDSG-----DGKF 698  
 Db 612 YVYQPGSDENDWTP-VMIYDSQVARKKTHYDGMGICIEBDDDGANGTSIGYGTI 670  
 QY 699 YTHIQGYDEQGRHHTSYSDYLTNGRQDTP-----DKVHLSMSKSYDNWQJANTHWS 753  
 Db 671 RKLVARQYDVLGQLSKESISNDWLMN--LSANPLVRLATPLVTTKTYKDGMLYSTEYS 728  
 QY 754 YGVSEKTYDPIITLTK-----QLQSNNSNVOTGKEVTTYPSQPIQITLFDAGHL 807  
 Db 729 DGEIELEIHPITRTITQGVKGLAMNIQNNF-----EOPASIKAVYPGTI 776  
 QY 808 QSCHTLTRDGMDEVRKETDAIGQCTTYQYDNVNRVIOITLPDGTIVRKXAPSTDTLIT 867  
 Db 777 YSTRITRYDGFGRVTTETDABGATQIGYDFDVIKTLPLDGTILSAYASFHEELIS 836  
 QY 868 DIRVNGISLQGFQFDGLSRLTOSODGRVAAVYYSAGNDQCPSTVTPPDQPIHYQYQPE 927  
 Db 837 ALNVNGQLGALVYDGGIRYISDTVGRKREYLYGPGDK-PIQSTIRSNKQMDLYY 895  
 QY 928 LDAVILQVANSNETTQOFSYNPVTGALIKAVAEQSLLPI-YRSGRLKMN-INDMKMS 985  
 Db 896 LGSVMSKFTTGTQONFRYSHKGTGLLSA--SEGVSGTNYVFPSPGVLAGRESFLADNRPIS 954  
 QY 986 ---YLTIRGLNGYDULNCTIKISPDTHGRVTOIKDSSIKTLNVDLNRHGSQVTD 1042  
 Db 955 SGELYTMSGLIRHNDOSFEHNHYVSVDQGRVLTKEQDQVAFEDVNDGRLLTTTKD 1014  
 QY 1043 LATGHMLTTFVBPEDGAINREIGRLCDSSGHTLIDQOSMLKTQOLANRIVKLVGLQRTQ 1102  
 Db 1015 TSLSLQVLTKEIDAPDREIKSLISDSIQV--ITLSYTNQNDISQITSIDGVMMNER 1073  
 QY 1103 YSYDSRRLNQYKQDAECPDYKSHSYTONFTYDIGNITACHTTPADGTEDHATFKF 1162  
 Db 1074 YQYDNNQRLSQYOCEGOSPIDHTGRVLANQIYHYQWGNIKRLDNTYRDGKET-VDYHF 1132  
 QY 1163 ANPTDQQLTEVHTHPMDMNRILKYDKAGRVINIT-DNHGNTENFTYTLGL---QN 1218  
 Db 1133 SQ-ADPTQLIRI---TSDKQQLSLSYDANG--NTRDEKQGT--LIYQNNRLVQVKD 1182  
 QY 1219 GQGSV---YGYDPLNRLVSQ--KTDTLDCELYRRETLVVEVNGEMIRLRTGETIIAQ 1273  
 Db 1183 RLGNWVSQYDMLNTLTAQVLANGVYNNQ-HYASGVTNMQUGDAELITLSSDKQIGH 1241  
 QY 1274 QRA-----SKVLLTGTSQOSVILTSQKUNLSQEAYSAYGKHSTANDASITLNGERADP 1329  
 Db 1242 QSAKNGQSVYYQYGIHNSSTVIASQENELMALSYTPYGRSLI---SSIPGLNQAQVDP 1298  
 QY 1330 VSGVTHLNGVRSYDVTLMREFTPDLSPPGAGINPYSCVCLDDPINRSPSGHLSMQAM 1389  
 Db 1299 VTQWYFLNGYRVFNPLMKRPHSPDMSPPRGGINPYTCQDPINRDILNGLHLSAGI 1358  
 QY 1390 TGIQMGJAGILLTIAT--GGMAIA---AAGGIAAIASTSTALAFGALSVTSDITSIV 1443  
 Db 1359 LGIVLAIIGIIVIGIVELGAGAAISAGLIAAGLAIASISALAAVATVIGLAADSIGIA 1418  
 QY 1444 SGALDASPKASSIILGWSVMGAGAGLALSALIKGKTKLATHGAFADGEGNALLKST--SE 1502  
 Db 1419 SALSSEKDPKTSIGILNMIISAGLGLVLSFGISAI-----TFT-----SSIVKSARSG 1463  
 QY 1503 SSRIKWGVTRSLDREIVRBEQVITDHSR 1532



Db 1007 LGVSPFIQOTSYLLPMGLHAEVGAQASAMD-RMWSGRAR-BETHDIRSGKKXKHYR 1064  
Qy 988 WTLRGLNGTYDLDLGTIOKISRDY--H-GAVTOIKDSITKTLTYNDLNRHISQVTDL 1043  
Db 1065 YSLTGNLEGADIDGAHARSYETAHAHVGLIADAAVTVTLAYDGLQRLCSWTARD- 1123  
Qy 1044 ATGNMLTTFYEPDGLNEIGRKLDDSGHTLIDIQSMKTOQLNRYKLVNGVLRTEQY 1103  
Db 1124 GRGHALTLTFEBSLGRHTKTLTAAESAEATLTSQEMVPCQLHQRKRSBGKPCDETF 1183  
Qy 1104 SYDSRNLMQYKCDGACBPTDKYGHSLYTONFTYDIYGNITACHTPADGTEHATPEFA 1163  
Db 1184 VYDARNLKOYAAAGPGPKAYGNALRGCKFEFDAPFNIRKCTTVADGSGENVGEVLF 1243  
Qy 1164 NPFDPCQLTEVHTHPD--MPDNIRLKDKAGRYINITDNHGTENFTYDTGL---QN 1218  
Db 1244 NPADPCQLTKVTNLSALDKGYPPAIEKLDQAGRLER--DEAG--RRLSYDALGRLARVEG 1299  
Qy 1219 GGGSV-YGYDPLNLVSGKTDY--LDCELYYRETMVNE--VRNG-----EMIRLLRT 1266  
Db 1300 GGGASVGYDAHDLVCCRVETSGMDHRLYRANRLVEMWTRSGQAPGADDDRVRLVYA 1359  
Qy 1267 GRTIAOOR---ASKVLNLTGDSQGSYILTSRKQNTSQEAYSAYGKHKT 1313  
Db 1360 AGSCAAVNEGDSVALMGTGDKGSIVSQAEGQAKHYATPYG-HQSS 1409

## RESULT 4

ABM67433  
ID ABM67433 standard; protein; 1787 AA.

AC ABM67433;

DT 20-NOV-2003 (first entry)

DE Photorhabdus luminescens protein sequence #530.

XX Anticbacterial; fungicide; insecticide; polymorphism; genetic analysis;  
KM detection; food; gene expression; plant; animal; microorganism; toxin;  
KM antibiotic; biopesticide; virulence factor; disease model; plague;  
XX whooping cough.

OS Photorhabdus luminescens.

XX KO200294867-A2.

PN 28-NOV-2002.

PF 07-FEB-2002; 2002KO-IB003040.

PR 07-FEB-2001; 2001FR-00001659.

PA (INSP ) INST PASTEUR.

PA (CNRS ) CNRS CENT NAT RECH SCI.

XX Duchaud E, Taourit S, Glaeser P, Frangeul L, Kunet F, Danchin A;  
PI Buchtleser C;

XX WPI; 2003-148459/14.

PT Genomic sequence of Photorhabdus luminescens and encoded polypeptides,  
XX useful e.g. as therapeutic antimicrobials and agricultural pesticides.

XX Claim 2; SEQ ID NO 530; 1205pp; French.

XX The invention relates to the isolation of genes and their encoded  
CC proteins from Photorhabdus luminescens. The isolated sequences are  
CC sources of probes and primers for detecting the genome of P. luminescens  
CC and related species; to study polymorphisms; for gene analysis and for  
CC detection/amplification of the genes. Antibodies (Ab) raised against the  
CC polypeptides encoded by the genes are used for detection/identification  
CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that

CC carry a gene-containing vector are used to select compounds that  
CC modulate, regulate, induce or inhibit expression of the genes in plants,  
CC animals or microorganisms other than P. luminescens and are able to alter  
CC response or sensitivity to toxins and antibiotics produced by P.  
CC luminescens. Cells transformed to express the genes are useful for  
CC recombinant production of the proteins, particularly toxins and  
CC anticbacterials useful as insecticides, bactericides and fungicides. The  
CC genes, proteins, vectors containing the genes and Ab are also useful  
CC therapeutically (to treat microbial infection by bacteria or fungi that  
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as  
CC biopesticides. Other uses of the genes and the proteins are as virulence  
CC factors and for identifying targets of human diseases for which P.  
CC luminescens is a model (particularly plague and whooping cough). This  
CC sequence represents one of the isolated P. luminescens proteins

XX Sequence 1787 AA;

Query Match 17.8%; Score 1583; DB 6; Length 1787;  
Best Local Similarity 28.2%; Pred. No. 4.1e-101;  
Matches 511; Conservative 280; Mismatches 727; Indels 294; Gaps 65;

Qy 22 TOANNFTSAVSGVDPRTGLYNIQITLGLVGNGLGPTLPLTSLSPKNTDIGFISGF 81

Db 30 SQATNFTGAMMGVDPRTGLFTPOIPYTGANDLWGPDLATLTYNPLTYLNTGYGLGF 89

Qy 82 NPGLSVDRKNSLSTSGENYKVIETDKY-----KLOQKLDNLRFEKDLKENCYRI 135

Db 90 SDNFTRYDTQVTLATYGBIYHREKANEVDQAVTFHRAKPAHKKVKE--KDAFW 147

Qy 136 IHKSGDIEVLTFGNNAFDLKVPRKLNPAHAIYIDMNFATOP-----RLNRIYDLDG 191

Db 148 LKITSREKLTOL--DRANPAVAVSEIYAPSGHLCVKNASFVNHNNTYQMLRECDAME- 205

Qy 192 HDIPLNLLEYQGLIKTLLPFGQKSGRYTELPRNLQMSIHNFSLGNENPLTWSEGYT 251

Db 206 ---TLKADLATEKEIEFTWPGSPESYTVLMTNDLLQTVISAS-----ELTWLLEYE 257

Qy 252 PIGKNGILGQWITSMRPAKGLKETYVSNNOGHHP----- 288

Db 258 TEGAHKKI---LKVTPPSGLIEVYV--HETGTLTPKCIOPYNPMGPIIRDPKS 312

Qy 289 -QSANLPLVPLPYTLMKQVPAGOPAIQAEYSYSHNVGSGSNGINV-----NKLDNLYGL 343

Db 313 STTRATVHFPPYVQHNLIAAGSBDQYRYVSEBNFLGQKNKMDPIPLPODNAVYLA 372

Qy 344 MTEYNVSTESRRYKQKGGHDIYRIBRYNNYHLTSECK-----QONGYIQTETAY 397

Db 373 NSEYKVTSTEVREYNEKR-----YCHREYNKPHLAVSETETVEVTPSRQKKEITIKY 427

Qy 398 YAIIGHNFD--SQSPQQLPKTETYSADNS-----YSEITETTFDSGNPLTYVIK 450

Db 428 YADVGSFDDNKPQPPMLPMTVEITWNHNESSASTORREKETTQWYNNQCNISMTL 487

Qy 451 DKKTOKIISPTMWEYYPAGEV---DNCPBPFGTFPFXKI--IOPPYD-SSEKDDPEK 505

Db 488 DNTTKT-----TYAPDGBETTTDHCAPBEGFERFIKEJAVEAPSPKTKITILIR 540

Qy 506 FIOYRYSLSQS-----HYTLKIEBHYS-----ATOL-----INSTLFQYNTD 545

Db 541 KTYNYKSYDTYSPKNNQASVVKSMVLPLSEHTHSRDCADHLEKVKINTVSYFENTQ 600

Qy 546 KSELGRLLKQTECKGNG-KTYSVNAKFTYTKODDTLQOQSHSITTHDN-----TIHRS 599

Db 601 NAFANGVEQRNSYVLFENGNQTRSYTEDEYSWNSNKNKAGCICKTKTSGCKGTPPSHE 660

Qy 600 QVRSRYGRLESPDPTFDIVTQMSYDKGLRLTRTNSGTPYANTLTYVDLNNLLODNR 659

Db 661 QVRSRYGRLESPDPTFDIVTQMSYDKGLRLTRTNSGTPYANTLTYVDLNNLLODNR 716

Qy 660 PPTVITTDVNGQLRNEFDGARRHSQCLKSDGDKFTYTHIQOYDEGRRHHTSYSD 719

Db 717 --VTVQTDIHERYITTEMGRLPRPKYSYPAGYQCKQDMERYQYNPLGQLCOALISCD 774

```

QY 720 YLTNGROOTDDBKVHLSMSKSYDNMGQANTHMSGVSEKITVDPITLTATKQLOSNNSN 779
DB 775 HML-FENKSEKCSMTVSLLEYDMGHNSRLRYSDGHSVKATPPIKMTSHQVSNDES 833
QY 780 VQTKREVITYTPSOQPIQITLFDAGHLQCHTLTRDGMDEVRKETDAIGQCTIYYQDNY 839
DB 834 QSSGIIRTYNQFOQITATERLTLSRIQOGCMHYLRDELGRLL-VISINANGWITLLAYDAF 892
QY 840 NRVIQITLPQGTIVNRKKA-PSFTDITLTDIRV--GISLGOQTFDGLSRITLQSDGGRV 896
DB 893 DRVVKQTFADDTTISMAVENGVSVMSATPIGINQTPILGTQILDGIRYIDMSEGRK 952
QY 897 WAYTYSAGNDQCFPVITPDGQ-----FHYQOPELDDAVYQVVS--NEITQOFSYN 947
DB 953 IKLDYE-GASFPVDTIVYVKPTSGDKDITVHYEPEPLNNAVTKITRAGDDVUQTFKTD 1011
QY 948 PVTGALLKAVAE-----GQSILPIYPSGRLKME-----NINDMKMSYLTWLR-G 992
DB 1012 PKTGLTTEALIEYKXNDKEMISKLIFDYTLSGOLASHKLVSYNNKNRAVYHFPKCYSPS 1071
QY 993 LENGTYDLTG-----IQKISRTHGRVQIQDSSKITLANTDULNRHIGSOVYDL-- 1043
DB 1072 PANPPTAKSTINGRXYIINKYSYDNQGRLETRNNDIEIMLKYDKLSRHVCSQSTEXLFD 1131
QY 1044 ATGHMLTTFVEFDDGLNREIGRK-----LCDSSGHT-LDIOQSWLKTQOLANRIVKLVNGV 1096
DB 1132 TINNKITTSILDLDPGRRETKTIRRHNGKNNKGTIDHISQTPREOKIKNNKL-LHGT 1190
QY 1097 LQ-RTEQYSYDSRNRLNQKCDGAECPYDKYGHSLVTVQ--NFYDIYGNITFACHTTFADG 1153
DB 1191 KQISKEVYTHYKRGGLFTYTL-----MELVGEREITQCSYKDKHLGNITQ--HSITFEG 1243
QY 1154 TEDHATFEFANPT--DPCQLTEVNH--THPDMEDNIRLKYDAGRYINITD--NHGNTENFT 1209
DB 1244 KITSTYYTGNGNIQDPCQLIDVSTCTGNSSSSLRFTYRNQALVCCENDNNKIRITWT 1303
QY 1210 YDTLGRLQNGQGSV-----YGYDPLNRLV--SOKTPT-LDCELYARETALVN----- 1253
DB 1304 YDSLRITTYDALHAKVETRIPLFATNMLIKRSKNGTPIHHDLSTYSSNSLVHDNRYG 1363
QY 1254 EVRNGEMIR-----LLRTGETTIAQ--QRAASKVLLTGTDSQOQSYILTSDKNTLSQE 1302
DB 1364 EKRDMAADRKNNKVGICLGFSPQTPCHQTPSVASRYETETATDQGSVIATFOGEDVQHI 1423
QY 1303 AYSAYGKRKSTANDASITIG-----YNGERADPVSGVTHLGNRYRSTDPITLM 1348
DB 1424 AYSFPG--VTTQAMVTVAGQOPPHNTAIEBPRFNGEQMDTASASYLLGNGVRAYRPDLM 1480
QY 1349 RFHTPDSLSPFGAGGINPYSCYCLGDPINRSDPSGHLISQAWTGMGIAGLLLTATGGM 1408
DB 1481 RFTAPDSMSPFGAGGINAYVCGSDPVNLNDPSGHISGMGANITTTGGIGILLAPFTYGG 1540
QY 1409 AIAAAGIAAIASTSTTALAFAALSVTSDITSIVSALGALDAPKASISILGVMSGMGA 1468
DB 1541 SLEIGLGVNAARGLL-----ALDAASGVTAISGALENKNPETSRLGMSGLG-- 1590
QY 1469 GLAESAITGCTKLA-----THLGAFA-----AEDGENALIKSTSE--- 1502
DB 1591 --LPSMVIGSYSLAQMNVNRLTNSPRTYHYPTSLGEVNLRSKSSDWMNARSINSGENW 1648
QY 1503 -----SSRIKKG--VTRSLDREIVRNEGOVIVDHSRGVTDNMGKGEQAI-----L 1547
DB 1649 HSEVGLNGRTTWGSDTKIRGLD--IKYPLEQISRRPSNG--DIYLLGSHGVQNGDWM 1703
QY 1548 VHGDQDGLYH-----TEGNKNGKGPYTRHTPRLQVLDYLDKNNIIVDLTQGDREV 1598
DB 1704 INGSRRGSLHHPFPKSDMTVYGSGWKGRITVHVR-----LATMSEIDPGLTLNNGNSHI 1758
QY 1599 HLLSTCYGKSSGA 1610
DB 1759 ILGYCYGRNDA 1770

```

```

RESULT 5
ABM67283
ID ABM67283 standard; protein; 1590 AA.
XX
AC ABM67283;
XX
DT 20-NOV-2003 (first entry)
XX
DE Photobacterium luminescens protein sequence #380.
XX
KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KW detection; food; gene expression; plant; animal; microorganism; toxin;
KW antibiotic; biopesticide; virulence factor; disease model; plague;
KW whooping cough.
XX
OS Photobacterium luminescens.
XX
PN W0200294867-A2.
XX
PD 28-NOV-2002.
XX
PF 07-FEB-2002; 2002WO-1B003040.
XX
PR 07-FEB-2001; 2001FR-00001659.
XX
PA (INSP ) INST PASTEUR.
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI Duchaud E, Taourit S, Glaeser P, Frangeul L, Kunet F, Danchin A;
PI Buchrieser C;
XX
DR WPI; 2003-148459/14.
XX
PT Genomic sequence of Photobacterium luminescens and encoded polypeptides,
PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX
PS Claim 2; SEQ ID NO 380; 1205pp; French.
XX
CC The invention relates to the isolation of genes and their encoded
CC proteins from Photobacterium luminescens. The isolated sequences are
CC sources of probes and primers for detecting the genome of P. luminescens
CC and related species; to study polymorphisms; for gene analysis and for
CC detection/amplification of the genes. Antibodies (Ab) raised against the
CC polypeptides encoded by the genes are used for detection/identification
CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
CC carry a gene-containing vector are used to select compounds that
CC modulate, regulate, induce or inhibit expression of the genes in plants,
CC animals or microorganisms other than P. luminescens and are able to alter
CC response or sensitivity to toxins and antibiotics produced by P.
CC luminescens. Cells transformed to express the genes are useful for
CC recombinant production of the proteins, particularly toxins and
CC antibacterials useful as insecticides, bactericides and fungicides. The
CC genes, proteins, vectors containing the genes and Ab are also useful
CC therapeutically (to treat microbial infection by bacteria or fungi that
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC biopesticides. Other uses of the genes and the proteins are as virulence
CC factors and for identifying targets of human diseases for which P.
CC luminescens is a model (particularly plague and whooping cough). This
CC sequence represents one of the isolated P. luminescens proteins
CC
XX
SQ Sequence 1590 AA:
17.7%; Score 1569; DB 6; Length 1590;
Best Local Similarity 30.8%; Pred. No. 3.3e-100; Indels 200; Gaps 57;
Matches 477; Conservative 236; Mismatches 636;
QY 63 LTLSSPLNKTDIDFGIGFNGFGLSVYDRKNSILSLSTGENYK---VIETDKTVKLQCKXL 119
DB 13 LNLSPFLTTIANFGIGWFRSLTMDVKTITFRSBNSEQKCKPLPNNNDISPKDKXL 72
QY 120 DNLRFEKDLKENCYRIIHKSGDIEVLTFGNNAADLVKVPKQLNPAGHAITYIDNFEATQ 179
DB 73 KDLRVYK-LDSNTFYVYVKNKGIIEFLKRIGSS--DIAKTVALFFPDG----- 116

```

QY 180 PRNRIYDLDGDHIDPLNLEYOGILKITILT--PPOKEGRTREIARL-NROLSIHNS 237  
 Db 117 ----EYFDLTYNRPALSEIKRMTGKTMYLKLNSGN---CTSVETPDNNLSAKIAFD 169  
 QY 238 LGNEPLTWSPGYTPICK-----NGILGOMITSMAPGLKETEYVNSNNOG 284  
 Db 170 YRNDYLITVVPDASPIDARPKMTYOTLKGI-F-VISAFTPTGYVLEVYKEN--G 226  
 QY 285 HMFPOSANLEFVLYTLMKOVPAQAPALQAEYSYTS-HNYVGGSGNGI-WNNKLDNLVY 342  
 Db 227 HKYTDIESIPYAALITIQ---PENGOPALISKSYEYSVHNPLFGLYSQRTSPDSQNLVYL 283  
 QY 343 LMTEYNYGSTESRRYKDKHEGHDQIVRERTYNNYHLLTSECQONGIQOTTERAYALIG 402  
 Db 284 VTGKYTYSSIE---RVLNGONVISTYERFDFKMLTKAKTQDNKRILITETLYNEDPS 339  
 QY 403 HNFDSOPSPQOLPK-----TKETWRSADNSYRSSETTETTPDSGNPLTKVICKKTKOK 456  
 Db 340 KSFSEBPENIQPSHYLTRITDLOQ-----NTSREBVNIKSDDMGNTLL-ITETSGIOK 393  
 QY 457 IISPTHEWYYPAGEVDCPPBPYGFTRFKKIIQTPYDEKRDDEKEFIQYRYSLIGS 516  
 Db 394 -----EYVYYPVNGEGNCPADPLGFSRFLKSVTQKSPDAQSVANRVTSYTYQKLP 447  
 QY 517 QSHVTLKIEERHSAQOLNSTL---FOY---NTDSELRLLKQTECTGSENGKTYSV 570  
 Db 448 FTGAAYK--EYVAKASSETISKIVRTFNYPNSPNNKH-GSLAKITYSVMMNQ-----QTV 499  
 QY 571 HKETTYKODDTLQOSSHITTHDNFTIHSQVRSRYTRLPSDPTDKOIVTOMSYDKLGR 630  
 Db 500 TTPKYEYSBEMTNTNVTVPFDGTHESKRVTSIYTHROLRKVDVNHVITDQYDLSGRI 559  
 QY 631 LTRFLNSGTPYANTLYTYVELNLIQDNRPVYITTTDVNGNQLRNEFDGAGHVSQCLK 690  
 Db 560 IGGIDPBTKEIKRSYIYQPGDENDPWP-VWIEIDSGIRKTHYDMGRICSIEBQ 618  
 QY 691 DSDG-----DQKFTIHTQYDEGRHHTSTYSYLTN-----GROOTDDKHLMS 737  
 Db 619 DDGCVMGTSIYQGYRKVLARQYDVLGQVKEISINDMWDLSANPLTRLITTP--LVTTK 676  
 QY 738 SKSYDNMGQJANTHWSYGVSEKITVDPITLATK-----QLOSNNSNNVGTKEVTTYTP 791  
 Db 677 TYQYDGMKNKYSTIEYSGRIEIEIHDPITITITQYVGLGMLNLIQNNP----- 725  
 QY 792 SQOPIQITLFEAGHLQSCHTLTRDGMWRVREKETAIGQCTIYOYDYNVNIQITLPDGT 851  
 Db 726 -EQPASIKVYVPGAIYSTRYVDGGRVTEITDAGVATQIEYDLFDRIYKKTLPDR 784  
 QY 852 IYURKTAPESTDILDIRVNGISLGQOTPDGSLRLTQSDGGRWAAVYTTSAQDQCPST 911  
 Db 785 ILSAAYASFHEBELISALNNGTQGLSYDGLGRVTRDVGKRTKEYLGLSGQDK-PIQ 843  
 QY 912 VITP--DGGFIHYQYQBELDAVLQVANSNEITQOFSNPTVGLALKVAEGOSILTP-YY 968  
 Db 844 SVTPANNKOIIDLK--ALGSVMSKPTTETSQONFSTIYQKTGLALSH-TEGVQSQNSYTF 900  
 QY 969 PSGRLEKMENTIN-DMKKS---YLMTLRGLNGYTLDTGTIOKISRTHGRVTOIKDSIK 1024  
 Db 901 PEGVLGHESFSDNKPFISSGDYRTYMSGLIORKHDSFAHDHVSYSABEGRLVTEBSSQY 960  
 QY 1025 TITANVDLNRHISQVTLDTLGHMLTTTVPFDGLNREIGRLCDSSGHTLIDIOOSMLKTQ 1084  
 Db 961 ATREYDNGVGLITTTKDTISLSQALAKIEYDVFDRKIKSLSDPSIQV-ITLSTYKNN 1019  
 QY 1085 QLANRIYKLVNGVLTORTQOYSDSRNRLNOYKCDGAECPCTKYGHSIYTOQFTYDIQNT 1144  
 Db 1020 QISQRTISIDGVKMKERYDSDNQRISQYQCEBEGSPVHTGRVLSQYIYHIDOMENIK 1079  
 QY 1145 ACHTTPADGEDHATKFPANPTPCQITLVNHTHPDMNDIRLKYDKAGVINIT--DNHG 1203  
 Db 1080 RLNDYADGKET--VDYHSQ-ADPTQIIRI-----TSDKQIEISLYDANG--NLTREDEK 1130

QY 1204 NTEFTYDYLGRL---ONGGSY---YGYDPLNRVLSQ--KTDITLDCELYYRETMVNEY 1255  
 Db 1131 QT--LTYDQNNRLVQYKDSKGNLVCOYQYDALNKLTAQVLANCTVRQ--YIASGVANYQ 1187  
 QY 1256 RNGEMIRLL-----RTGETIIAQORASKVLLTGTSQOSVILTSRKONLSQEA 1303  
 Db 1188 LGDETITWLSQKRLGHQSTKNKESVYYQ-----GTDHNSYVIASQNEIEMALS 1239  
 QY 1304 YSAVGHKSTANDASILGYNGERADVPYGVTHLGNCRYSVDPTLMFPHRPDISLPPGAGG 1363  
 Db 1240 YTPYGRSLI---SSLPNGNGAQVDPVTQWYFIGNCYRVFNPLMFPSPDSWSPEGRGG 1296  
 QY 1364 INPYSYCLGDPINRSPSGHLSQAWTGIQMGJAGLLTIAI--GGMAIA---AAGGIA 1417  
 Db 1297 VNPYTYCQGPINRIDLNGHLSAGILGIVLGAIGIIVGVVSIGAGALISAGILIAAGAL 1356  
 QY 1418 AAIASSTTALAFALSVTSIDTISYVSGALBEDAPASSILGVNSMGMAAGLAESAI-- 1475  
 Db 1357 GAIASISAPAVATATVIGLADSIGIASALSEKDPITAGILNISTGVLSPGISAITF 1416  
 QY 1476 -----KQGTKLATH--LQAF-AEDGENALIKSTSESRITKWGYTRS 1513  
 Db 1417 TSSLIKSARSGSQVASTSVIGSVPIEFGEIA-----SRSSR-RMDIALS 1459  
 RESULT 6  
 ADG73113  
 ID ADG73113 standard; protein; 982 AA.  
 XX  
 AC ADG73113;  
 XX  
 DT 11-MAR-2004 (first entry)  
 XX  
 DE Pseudomonas syringae pv. tomato DC3000 Avr/Hop protein #53.  
 XX  
 KM Avr; Hop; transgenic plant; disease resistance; cancer; bacteria;  
 XX metabolic pathway; eukaryotic cell death; programmed cell death;  
 XX cytosstatic.  
 OS Pseudomonas syringae; pv. tomato str. DC3000.  
 XX  
 PN US2003204868-AI.  
 XX  
 PD 30-OCT-2003.  
 XX  
 PF 12-FEB-2003; 2003US-00365742.  
 XX  
 PR 12-FEB-2002; 2002US-0356408P.  
 XX  
 PR 10-MAY-2002; 2002US-0380185P.  
 XX  
 PA (COLI/) COLIMER A.  
 XX (ALFA/) ALFANO J R.  
 XX (CARF/) CARTINHOUR S W.  
 XX (SCHN/) SCHNEIDER D J.  
 XX (TANG/) TANG X.  
 XX  
 PI Collier A, Alfano JR, Cartinhour SW, Schneider DJ, Tang X;  
 XX WPI; 2003-875735/81.  
 XX N-PDSB; ADG73112.  
 XX  
 PT New nucleic acid, useful in imparting disease resistance to a plant or in  
 XX preparing a composition for treating cancer.  
 XX  
 PS Claim 15; SEQ ID NO 106; 209bp; English.  
 XX  
 CC The present invention relates to the isolation of Pseudomonas syringae  
 CC pv. tomato DC3000 Avr/Hop proteins, and the polynucleotide sequences  
 CC encoding them. Also disclosed are expression vectors, host cells, and  
 CC transgenic plants comprising polynucleotide sequences of the invention.  
 CC The polynucleotide and polypeptide sequences are useful in imparting  
 CC disease resistance to a plant or in preparing a composition for treating  
 CC cancer. The sequences may also be used to make a plant hypersusceptible

to colonisation by nonpathogenic bacteria, modify a metabolic pathway in a cell, cause eukaryotic cell death, and inhibit programmed cell death.  
 CC The present sequence represents a *Pseudomonas syringae* pv. tomato DC3000  
 CC Avr/Hop protein of the invention.

XX Sequence 982 AA;

Query Match 10.5%; Score 932.5; DB 7; Length 982;

Best Local Similarity 29.7%; Pred. No. 8.3e-56; Indels 175; Gaps 33;  
 Matches 285; Conservative 142; Mismatches 357;

```

QY 642 ANTLLTYEELNNLODDNRPPFVITTTVDVNGNQLNEFDGAGRHSQCLKSDGDKETTYI 701
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 31 SSTLGKRYD-----DWMQR--CCTTTDDN-VQTYEYSDPIGSDVHK-----GPIQKT 74
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 702 HTQGYDEQGRHHSTYSYDLTNGRQCTDPDKVHLSMSKSYDNWQQLANTHWSYVSEKIT 761
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 75 WKQSGDEGR-----ISGRSET-----WNLN-----FGKPDPIR 103
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 762 VDPITLTATQLOSNNNVQTKGKVTYTPSQPIQITLPDEAGHLQSCHTLTREDGMDRV 821
   |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 104 ----TLTAGKTKGRKRTISMRSRNLTT---TEQEL-----SRQTFLYDGLGRC 144
   |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 822 RKETDAIGQCTIYOYDNNVNIQITLPDGTIVNRKVAFFSTDTLITDIRV---NGIS--- 875
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 145 TEGRDALQOSTLPSYDWMRSWVSSTLADGVIYNSYAPQSSSELATMLEVHNGTTRTV 204
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 876 LGQOTFPGLSRLTOSQDGRWMTYTSAGNDQCSYITPFGQFIHQYQBELDDAVLQV 935
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 205 AGTKFPGLEKRVYQTKGDRVEQFNVDAGEMQ--PRSRTAGLDNINFTYTRALTDQIFSS 263
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 936 ASNEITQOFSYNPPTGALLKVAEGQSLTPYYPGSRKLMENINDMKMSYLM-----T 989
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 264 TAPDETAKFYDXTKSARLIEATNQGRTTYVDVHNLQGTWNL--LGAMETRRQSS 321
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 990 LRGLNGEYTDL----TGTQKISRDTHGRTVQIKDSIKITLTYNDLNRIHISQVTDLA 1044
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 322 LLGSRPKRTDLKKGAGAEATRYDYDLGIRFINQSLRTIIDYDVLGQCKVATEDLQ 381
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 1045 TGHMLTTFVEFDGLNREIGRLCDBSSGHTLDIQSMKLTQOLANRYKLVNLTQRTQYS 1104
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 382 AGTVIITDMEYDDQGEITLKTQTALSNQALTLTQTAWDGLKTRDLOQASSPLIHETFS 441
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 1105 YDSNRRLNQYCKDGAECPTDKXGSIYQNTFYDIGNITACHTTFADGETDHAETFEAN 1164
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 442 YDPKRLTLVNYLSSLPREDLQREMTQIFSPBELDNITLCQRTFTDGTSEBAAFKGS 501
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 1165 P-----TDPQQLTEVNAHTHPMPDNIRLKYPKAGRVINITNHNANTENFTYDLAGRL-- 1216
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 502 PGDDKHKRCQLLSIAVTPPRKTPDPFSDYDANGN--QLKDBHGN--SLHYDSQSLRLQV 557
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 1217 -QNGOG--SVYGYPRNLNLYSOKTDTLDCEL--YTRFMLVNEVRANGMILLRTGETII 1271
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 558 AETGAPISQRYRYGHNQLVATR--DGNSEILRFEGHQLSSTVQEDERTQYIHLGEPPL 616
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 1272 AQQ--RASVYLCTGDSQSVILTSDKONLSOAYSAVYGHKKSSTANDASILIYNGERAD 1328
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 617 GQQIVDDAEQTLILLITDANQSVMEGFQOGQLRKAVYSYGERHSSEALLSTAGRNGEVRE 676
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 1329 FVSGVTHLNGYRSYDPTLMRPHRTDLSLSPGAGGINYSYCLGDPINRSDPSGH----- 1383
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 677 AANGVYLLNGGYRAVNPILMRPHSDPLSPFAEGGVNPFYCLGNPILALRDPPTHGDSAGQ 736
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 1384 -----LSNQ-----AWTIGMGIA-----GLLITATGMAIAAAGI----- 1416
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 737 TGRLRPDEGALPMQGGGDMGWVGIGVFTVLGVAATITATLGTPTVTPGPTVLGI 796
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 1417 -----AAAIASSTTALAFG-----ALSTSDITSIVGALGEDASPKASSILGWVS----- 1462
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 797 SMTASAAAIVSTVSGALIVGTALTAASTANTAVLVNN-----DQTGEGVGMGLGIAAV 851
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 1463 ---MGMAAGIAESAIKGTGLA-----THLGAFAEDGENALLKSTSESSR 1505
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

DB 852 PVGLVGFAGAVVAVAAAAKVAANAAGTIGVRSYRIG-LAAAGARRTISASASAR 909

RESULT 7

ADL12168  
 ID ADL12168 standard; protein; 982 AA.

AC ADL12168;

DT 06-MAY-2004 (first entry)

DE *Pseudomonas syringae* anti-cancer protein #53.

KM cytostatic; gene therapy; Avr; Hop; cancer.

OS *Pseudomonas syringae*, pv tomato DC3000.

PN W02003068930-A2.

PD 21-AUG-2003.

PF 12-FEB-2003; 2003WO-US004450.

PR 12-FEB-2002; 2002US-0356408P.

PR 10-MAY-2002; 2002US-0380185P.

XX (CORR ) CORNELL RES FOUND INC.

PA (USDA ) US SEC OF AGRIC.

PA (UTNE-) UNIV NEBRASKA.

PA (UNIV ) UNIV KANSAS STATE RES FOUND.

PI Collmer A, Alfano JR, Cartinhour SW, Schneider DJ, Tang X;

XX WPI: 2003-679632/64.

DR N-PSDB; ADL12167.

XX New nucleic acid molecule, useful for preparing a composition for treating cancer.

XX Claim 15; SEQ ID NO 106; 284bp; English.

XX The invention relates to novel *Pseudomonas* Avr and Hop genes, a sequence that hybridizes with these sequences under stringency conditions

CC comprising a hybridization medium that includes 0.9 x saline sodium citrate (SSC) buffer at a temperature of 42 deg C. The nucleic acid

CC molecule is useful for preparing a composition for treating cancer. This sequence corresponds to one of the proteins of the invention.

XX Sequence 982 AA;

Query Match 10.5%; Score 932.5; DB 7; Length 982;

Best Local Similarity 29.7%; Pred. No. 8.3e-56; Indels 175; Gaps 33;  
 Matches 285; Conservative 142; Mismatches 357;

```

QY 642 ANTLLTYEELNNLODDNRPPFVITTTVDVNGNQLNEFDGAGRHSQCLKSDGDKETTYI 701
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 31 SSTLGKRYD-----DWMQR--CCTTTDDN-VQTYEYSDPIGSDVHK-----GPIQKT 74
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 702 HTQGYDEQGRHHSTYSYDLTNGRQCTDPDKVHLSMSKSYDNWQQLANTHWSYVSEKIT 761
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 75 WKQSGDEGR-----ISGRSET-----WNLN-----FGKPDPIR 103
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 762 VDPITLTATQLOSNNNVQTKGKVTYTPSQPIQITLPDEAGHLQSCHTLTREDGMDRV 821
   |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 104 ----TLTAGKTKGRKRTISMRSRNLTT---TEQEL-----SRQTFLYDGLGRC 144
   |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 822 RKETDAIGQCTIYOYDNNVNIQITLPDGTIVNRKVAFFSTDTLITDIRV---NGIS--- 875
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 145 TEGRDALQOSTLPSYDWMRSWVSSTLADGVIYNSYAPQSSSELATMLEVHNGTTRTV 204
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 876 LGQOTFPGLSRLTOSQDGRWMTYTSAGNDQCSYITPFGQFIHQYQBELDDAVLQV 935
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 205 AGTKFPGLEKRVYQTKGDRVEQFNVDAGEMQ--PRSRTAGLDNINFTYTRALTDQIFSS 263
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

QY 936 ASNEITQFSEINPVYTGALTKAVAGSGLTPITYPSGRLKMENTINDMKMSYLM-----T 989  
 DB 264 TADERTAKFYDKTSNAILERINPQGRTRTRYDVHNOJTGSETMDNL--LQAMETHRQSS 321  
 QY 990 LRGLNGVYDL-----TGTIOKISRDPHGVTOIKOSIKITLTYNDLNLHIGSQVTDLA 1044  
 DB 322 LLAGRPKRTDILKKGGAAGAEFRYDYLGRIRFINGNLMTITDIDYDLGGLCKVATBDLQ 381  
 QY 1045 TGHMLTTVEPFDGLNREIKRLCDSSGHTLIDQOSMLKTQOLANRYKLVNGVLORTQOYS 1104  
 DB 382 AGGVIIIDMEYDQOGGELIRTOGASNOALTLTQYMAVDGLKTRDLOQAGSPULHETFS 441  
 QY 1105 YDSRNRNLRQKCDGACPTDKGHSIYTONFTYDIGNITACHTTFADGDEHATFEKAN 1164  
 DB 442 YDRGRGLTVLVNYGLSSLPDELOREMTROIFSPEDLNITLCOTRFDTGSERAAFKYS 501  
 QY 1165 P-----TDPQOLTEVHTHPDMPDNIRLKYDKAGRVINITDMNGTENFTYDTLGRU-- 1216  
 DB 502 PGDDKHKDRQQLSIATTPRKTPDPFYSVDANGN--QLNDEHGN--SLHYDSQSLLOV 557  
 QY 1217 -QNGQG--SYGVDPLNRLYSQKTDTLDCEL--YRETMVLVNEVRNGEMIRLRTGTTI 1271  
 DB 558 AETGGAFLISQRYDGNHQLVATR-DGENESIRREFEGHQLSSTQEDQRTQYLHGEPL 616  
 QY 1272 AQC---RASVYLLTGTDSQSVILITSDKONLSQASVAGKHKSTANDASILGYNBRAD 1328  
 DB 617 GQQLVDDAEQTLILLITDANOSVMGEFOGQURKAVISAYGRHSERALLSTAGNGSVRE 676  
 QY 1329 PVSQVTHLNGVRSYDPTLNRFTPLSPFAGAGINPVSYCGLDPIRNSDPSCG----- 1383  
 DB 677 AANGWYLLNGYRAYNPLNRHSPDLSPFABEGVAPYCYGNPILALDPTCHDSGQ 736  
 QY 1384 -----LSWQ-----AMTIGMGIA---GLLITATGMAIAAGI----- 1416  
 DB 737 TGRLRPEDEGALPMQCGGDMGWGIGIVFTLVAAITATLGTATVTVGTYVLGI 796  
 QY 1417 -----AAAIASSTTALAFG---ALSVTSDITSYSGALBEDSPASSIILGVNS----- 1462  
 DB 797 SMTASAAAVSTVSTGALIVGTALTAASTANTVAIVNN-----DQTAGEVGGMLGIAAV 851  
 QY 1463 ----MGWGAAGLAESAIKGGTKLA-----THLGAFAEDGENALLKSTSESSR 1505  
 DB 852 PVLGVFGAGAVVAVRAVAAAKVAANAAGTIGVRSVGRIG-LAAAGARRITSSAASGAR 909

RESULT 8  
 ADS22210 standard; proteoln; 1317 AA.  
 XX  
 AC ADS22210;  
 DT 02-DEC-2004 (first entry)  
 XX  
 DE Bacterial polypeptide #11243.  
 XX  
 KM Recombinant DNA construct; transformed plant; improved plant property;  
 KM cold tolerance; heat tolerance; drought tolerance; herbicide; osmotic;  
 KM pathogen tolerance; pest tolerance; plant disease resistance;  
 KM cell cycle pathway modification; plant growth regulator;  
 KM homologous recombination; seed oil yield; protein yield; carbohydrate;  
 KM nitrogen; phosphorus; photosynthesis; lignin; galactomannan;  
 KM bacterial polypeptide.  
 XX  
 OS Bacteria.  
 XX  
 PN US2003233675-A1.  
 XX  
 PD 18-DEC-2003.  
 XX  
 PF 20-FEB-2003; 2003US-00369493.  
 XX  
 PR 21-FEB-2002; 2002US-0360039P.

XX  
 PA (CAOY/) CAO Y.  
 PA (HINK/) HINKLE G J.  
 PA (SLAT/) SLATER S C.  
 PA (CHEN/) CHEN X.  
 PA (GOLD/) GOLDMAN B S.  
 PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;  
 DR WPI, 2004-061375/06.  
 XX  
 PT New recombinant DNA construct comprising a promoter positioned to provide  
 PT for expression of a polynucleotide encoding a polypeptide from a  
 PT microbial source, useful for producing plants with improved properties.  
 PS Claim 1, SEQ ID NO 11243; 122pp; English.  
 XX  
 CC The invention relates to a recombinant DNA construct comprising a  
 CC promoter functional in a plant cell, where the promoter is positioned to  
 CC provide for expression of a polynucleotide encoding a polypeptide from a  
 CC microbial source. The invention also relates to a transformed plant  
 CC comprising the recombinant DNA construct and a method of producing a  
 CC transformed plant having an improved property. The plant is a crop plant  
 CC such as maize or soybean. The method of producing a transformed plant  
 CC having an improved property comprises transforming a plant with the  
 CC recombinant DNA construct and growing the transformed plant, where the  
 CC polynucleotide or polypeptide is useful for improving plant properties.  
 CC The recombinant DNA construct is useful for producing plants with  
 CC improved plant properties, e.g. improved cold, heat or drought tolerance,  
 CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,  
 CC increased resistance to plant disease, better growth rate by modification  
 CC of the cell cycle pathway with plant growth regulators, increased rate of  
 CC homologous recombination, modified seed oil or protein yield and/or  
 CC content, improved yield by modification of carbohydrate, nitrogen or  
 CC phosphorus use and/or uptake, by modification of photosynthesis or by  
 CC providing improved plant growth and development under at least one stress  
 CC condition, improved lignin production or improved galactomannan  
 CC production. This sequence represents a bacterial polypeptide used in the  
 CC scope of the invention. Note: The sequence data for this patent did not  
 CC form part of the printed specification but was obtained in electronic  
 CC format from USPTO at [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html).  
 XX  
 SQ Sequence 1317 AA;  
 QY  
 Query Match 5.8%; Score 517.5; DB 8; Length 1317;  
 Best Local Similarity 21.2%; Pred. No. 2e-26;  
 Matches 334; Conservative 194; Mismatches 505; Indels 541; Gaps 72;  
 QY 39 TELYNIQITLGHIVNGNIGPTLPLTL--SYSPLNKTDIGFGIG--FNGELSYVDRKNSL 94  
 DB 56 TGSYFYQYODLSIPKRG-----LPLTVSRSYNMDNRSGIFGSGWTFNVMNLTVDNNGN 110  
 QY 95 LSLSTGENYK---VIEFDKTVKLQOKKLDNLREKDLKENCYRIHKSQDEIVLTGFNNN 151  
 DB 111 VTVLGGDGHDTDTIAMPDGYRSPLSVFDDL-----INKSDBYTLTKKDDQT 157  
 QY 152 AFDLKVPKKLLN--PAGHAIYDMNPATOPFLNRIYDDLGDHDIPLNLAYQGLIKTI 208  
 DB 158 KYNFSKGLKVINIVDKNGNQI---NFTYGEQLTVY-TPASGREL-ILAYDHNG----- 206  
 QY 209 LTLFPQKSGYRRLFLNRQLNSIHNFSLGNENPLTWGSGYRPIG-----KNGILGQ-- 261  
 DB 207 -----HISTIDMGKRWSTSYDDQGLLICKNPIGGKLS 241  
 QY 262 -----WTSMTAPG---LKETVYVSNNOGHFPOSANLPLVPLPYTLMKQVPGAGOP 311  
 DB 242 YTYDENHMTSITDPRGNHPMKNTY---DEKHVLSQS----- 276  
 QY 312 AIOAERYSYSHNVYGGSGNINNNKLDNLGLMTBYNYGSTRRYKXGEGHQIVRIER 371  
 DB 277 -----NSILATY---TENYDS-ENRK----- 293  
 QY 372 TYNNVHLLTSECKQNGYIQTETAYYAIIGHNFDQSPQFOLPKTKTETWRSDNSYRS 431



```

Db 294 -----TTFD-----PFGNKTYSPBHW-- 313
Qy 432 ELET-----TFDESNPLTKVIKDKTKOIIISPSTHWEYYPAGEVDCPPPYG 482
Db 314 ELNETNOGYTISAYDENGRIISVTNENSKTYLA-----YDANGNIK-TTNPBG 364
Qy 483 FTRPVKKIQTTPYSEFKDEBEKIQRYSLISQSHVTLKIEBRHSAQOLNS----- 537
Db 365 YSK-----SMTYDS-----KNLISQDELRHKTSPEDNSNLKISDALG 406
Qy 538 --TFQYNTDKSELGRLLKOTECTK-----GENGK-----TYSVHKFT 574
Db 407 HEYFVSID-----KGYOIGETDSNKKTATPESYNNNGOITTTDANGTSATFYTVGRVT 462
Qy 575 -----YTKQDPTLQOSHSTTHDNFTIHSQVSRVYGRJ-----FSDPTKDI 618
Db 463 TKTDAGKNRYTFQYDALDNLISID-----PMGQTTSTNTMLLEIKVLQMLKVXQSI 516
Qy 619 VTQMSYDKLGRLLRTIANSCT--PYA-----NTLYDYELANLQDDNRPPPVIT 665
Db 517 LTTIXISXIKKRMQWVEXLGTNNMPLAIWPPROTQKHKTISYDDPLNRQ-----VS 568
Qy 666 TTDVNGQLRNEFPDAGRHVS-----QCLKDSGDKFYTITHTQOYDE 708
Db 569 VTNALGKTRRKVYALGNKISITNAYGKSTRYSNSLNQVKTVMNGK--VVRNYDA 625
Qy 709 QGRHHTSYSD-----YLTNGRQOTDPRKVLHSMKSYDNWQOLANTHWSYVSEKIT 761
Db 626 VGNLISITDENGHKINGYDNLNQVSWTDALRKTTRNKYDANGNKISITNAYGKSTRYS 685
Qy 762 VDPITLVRATQLOSNMNVQGEVTTYTPSQPIQITLPEAGHLTOSCHTLTMDGMDRV 821
Db 686 Y-----NSLN-----QLVKYTDAMGQVRYNYDAVGNLIS----- 715
Qy 822 RKETDAIGQCTIYOYDNYNRVIOITLPGDTIVNRKYPFSTDTLITDIRVNGISLGQOTF 881
Db 716 --TTDANGRKTNVGYDSLNRQVSTITNALGKTRRKVYDANGNKISSTOANMR--LTKKSY 770
Qy 882 DGLSRLTOSQDG-GRWAVYTYSA-GNDCCPSTVITPD--GQFIHQY-----QELDDAV 932
Db 771 DSLRLVYKVTAMGQVRYTYDAVGN--LISTTDAKGKHTDGYDLSLRQVSIIDPL 825
Qy 933 LOVASNETTQGFSPNVPYGALLKAVBEGOSLPIYIPYSGRLKMNINDMKMSYVLTMLRG 992
Db 826 GRTLRNK-----YDANGNKISSTDEDEKITSYGVDLNRLTKYSYDDQVSY----- 873
Qy 993 LENGVTLDTGTLOKISRPDTHGRVYTOIKDSSIKTTLNYDNLNRH-----GSQV--TDLA 1044
Db 874 -----NYDAVGNRLTMKDSHGTTAKYDKLNRLLSVLNPDAQVSYTYNK 918
Qy 1045 TGHML-----TTVEFDGLNREIGRLCDSSEHTL---IQOSWTKTOQLANRYKL 1093
Db 919 VGNRKMTYTPDGKTTYSYDAVNRLIG--VIDSDHITSYSYAKNGNKTKMTNPGV--- 973
Qy 1094 NGVLORTQGYSDSNRLNOYKCCDGAECPTDKYGHISYQNPYDIYINATACHTTTFADG 1153
Db 974 -----KTB-YSYDKANRL-----VELINKTTOVASSYKTTLDAAGNRLKVDQLAEG 1020
Qy 1154 TED-----HATFKFANPDPQCLTEVHHTHPMPDN--TRLKYDAGRANIT- 1199
Db 1021 VESGDSLEKESQLLTTTYGY--DKLYRLTKV--DYPNKTYVSXKYSKDSGNRISMTT 1072
Qy 1200 --DNHGTENTFYDTLGRLONGQGSVGYDPLNRLVSOQTDTLDELY-YEFTMLVNEVR 1256
Db 1073 NVDGISTISYKYDADQLOSGNISYSYDKKGNLIKRVANSJQPMSPSYSE--ANLX 1129
Qy 1257 N-----GEMIRLRLT---GFTIIAQORASVULLTGDSDQOSVILT 1293
Db 1130 NVSEFVSNTNTPKSSYNFEPYDGNRIITKTITNGE--NAOSTKYVLDINSALPQVLT 1185
Qy 1294 SDKQNL-----SOEA-----YSAYGKHSSTANDASIL----- 1320

```

```

Db 1186 SDTKNTTCYTYGTDILISMTNSENAEYHYHGLSVRSLSDSKGIKNTYLYDAFGQVCK 1245
Qy 1321 -----GYNERADPVSGVTHLNGVRSYDPTLMRFTTPSLs--PRGAGINPYSY 1369
Db 1246 EIGTVNBEFFTEGQWDEGLYLRARY--YDPSVGRFTTKOVIGRRVTTQSINRIVY 1303
Qy 1370 CLGDPINRSDPSGH 1383
Db 1304 TTNNPVNLVDLTGY 1317

RESULT 9
ABG31849
ID ABG31849 standard; protein; 2334 AA.
XX
AC ABG31849;
XX
DT 05-NOV-2002 (first entry)
XX
DE Human kinase, MEK1.
XX
KW Extracellular signal regulated kinase; hyperalgesia; surgery;
KM opioid withdrawal; pain sensitization; analgesic; chronic pain; ERK;
XX MEK1; human; enzyme.
XX
OS Homo sapiens.
XX
PN W0200258687-A2.
XX
PD 01-AUG-2002.
XX
PE 25-JAN-2002; 2002WC-US002128.
XX
PR 25-JAN-2001; 2001US-0264336P.
XX
PA (TEXA ) UNIV TEXAS SYSTEM.
XX
PI Gusterin HB;
XX
DR WPI; 2002-608414/65.
XX
DR N-PSDB; ABK30804.
XX
PT Reducing or reversing tolerance, physical dependence, hyperalgesia,
PT withdrawal symptoms, or pain sensitization in patients on analgesics for
PT chronic pain, comprises inhibition of the extracellular signal-regulated
PT kinase (ERK).
XX
PS Disclosure; Page 156-161; 163pp; English.
XX
XX
XX The invention relates to a method of reducing or reversing tolerance,
XX reducing the risk of physical dependence or hyperalgesia, reducing the
XX symptoms of opioid withdrawal or inhibiting pain sensitization in a
XX patient taking analgesics. The method comprises administering an
XX analgesic and an extracellular signal-regulated kinase (ERK) inhibitor
XX comprised in a formulation to reduce or reverse tolerance, risk of
XX physical dependence, hyperalgesia, symptoms of opioid withdrawal, or
XX inhibiting pain sensitization in patients taking analgesics for chronic
XX pain or those undergoing surgery. The present sequence represents the
XX amino acid sequence of human MEK1 (not defined)
XX
XX
XX Sequence 2334 AA;
XX
XX
XX Query Match 4.9%; Score 432; DB 5; Length 2334;
XX Best Local Similarity 20.4%; Pred. No. 5.2e-20; Indels 560; Gaps 80;
XX Matches 358; Conservative 226; Mismatches 613;
Qy 44 IQITLGHIVG-NGNL-----GPTLPLTISYSPINKTIDIGFGIGNFGLSVYDR 90
Db 896 IDIPSGQLNATGAVINNEBDLSIDRGPGGLSRTYNSLSDSHLFGQGW-----YADA 950
Qy 91 KNSILSLSTGENYVYETDKTVKLQOKKLDN-----LRFEDKLKENCRTRIHKSQDIEVL 145
Db 951 ETSVISTDGMAY--IDEDATTRRFTTKKADGTYOPPTGVVLELTERADQFIKTKD----- 1004

```



```

QY 146 TGFNNAFDLKVPRKLL-----NPAGAIYIDMN-----FPAQPRLNRIYDDLOGH-- 192
D 1005 ---QTMAYFNKKGKLOKVVDDGHNNAATVYVNDKQULAITDASGRKLTTPYB--NGHVT 1060
QY 193 -----DIPLANLEY--OGEL-----KTLITLTPG 214
D 1061 SINGPKKAVTSYENDLKKVTDGTVSYDSDSGLRVKQYSA NSTAKVFEY-- 1118
QY 215 QKGYRTFLRLKQLNSIHNF-----LGNENPLTMSFGTTPPGKNGILGOWITSM 267
D 1119 QYSGHRLERKAINAKETEVVSYADAKKTLTMTQPNGRKVOGYVEANP-----ICVID 1172
QY 268 APGLKKTVA--YNNN-----QGHFPOSANLPVLYTLMKQVPGAG 309
D 1173 DABGLKTTTKTEGNNVVEDVDNDVGTGKATESYQYDKGN-----VTSVADAVGT- 1225
QY 310 QPAIOAEVSYSHNVYVGGSGNGIMNNKLDNLGYLMEYVNGSTESRRYKDEGHDIQVRI 369
D 1226 -----ETVEYNKNDV-----TKMKDEGNVTDIAYDGLDAVSETDQSGKSSAAV 1271
QY 370 ERTYNNYHLLTSECKQONGYIQTETAYVAIIIGNPDSQPSQFOLPFTKTEWRSADNSY 429
D 1272 YDKXGNOIGSSKDISASTNLIK-----DSFEAKQSGWMLTASKO----- 1311
QY 430 RSEITETTFDESGNPLTKVKKDKTKKILSPS-----THMEYPPAGVNDGCPPEPVFTR 485
D 1312 RRRKIS-VIADKSG-----VLSSGKALEVLSOSTSAGTDHGISATQVLEBPNTTYLSC 1365
QY 486 FVKKII---QTPYSEFKDDEPEFIQY---RYSLIGSOHVTLKIEBRHYSATOLLNS-- 537
D 1366 KITDILAKSARVAFIDARDQKRIOWIHNEYSLAKNDWT-----KROITFTTPANAK 1421
QY 538 ---TLFOYNTDKSELGRL-LKQTECTYGENKTYSVNHKFTYTKODDTLOOS--HSITTH 592
D 1422 AVVYMEVDHDKDKGKAMFDEVQLEKEGVSSYNPQNSSFTSATENNVVSGASVDSSE 1481
QY 593 NPTIHNSQVRSRTGRJFSDDTDFKDIYTKMSYDLGSLTFT-----L 635
D 1482 GFNDVSVLKARATISASQASVTQTVLQGSANDKPVYLTITLTKMSKASVKTDEKDYSL 1541
QY 636 NSGTPYANTLTYDEL---NNLQDDNRPPIV----- 664
D 1542 QANVTYADSGTGYNAKFPBGTOEMRAAVIIPTRKINKVDISILFQKSA TGVWDDI 1601
QY 665 -----TTVDVNGNQLRNEFDAGRHVSQCLKSDGKFTYTIHQYDEQGHNT 714
D 1602 RLIEGSLTKSTYDSNGNYTKEDDELGVATS---TDYDETKG---KTSETDAKGEKTY 1654
QY 715 STY---SPYLTNGRQOTDPDKVHLSMSKSYDNWG-QIANT-----HHSYGVSEKI-- 760
D 1655 YTYDQAQLTNMTLSNGSTSLH---SYDKENGEVSKTRAGADQTYKREYDVMGLVK 1709
QY 761 TVDPITLTATKOLQSNENNVT---GKEVTTYTPSQOPIQITLFDABGHLSCHTLTRD 816
D 1710 TTBPGLGNVLASBYDANSNLTKTISPNGENV-----SLSYD 1744
QY 817 GMDRVKRETDALGQCTIYQYDNVNRVITQLPDGTIVNRKXAPSTDTLTITDIRVNGISL 876
D 1745 GTDRVKSYSYNGTEKYLFTYDKNGN-----ETSVNKEGN-----TT 1781
QY 877 GQOTFQDLSLTLQSGDGRWATYYSAGNDQCESTVTPBQOPIHYQYBELDAVLAQVA 936
D 1782 KKRTFDDKNNLTETLDRGSGQWTWYPPDSDLKTF-----SWH-----G 1821
QY 937 SNETIQOFSYNPVYGALLKAVAGOSLTPYYPBSGRLMENINDMKMSYMLTLRGLENG 996
D 1822 DQKGTNQPTN-----KIQOMTEMKSTSSYSDYEN- 1854
QY 997 YTDITGTIQKISRDTHERVTOIKDSSIKTTLANDLNRHIGSQVTDLATGSHMLTTTVFED 1056
D 1855 -----GNVQ-----FTITNGGCTSPSYDERMLVSLIHGDKNGDILTESYEV- 1898

```

```

QY 1057 GLNREIGRKLCSSGHTLIDQOSMLTKQOLANRIVLNGVLQREQYSYDSRNLNQYKC 1116
D 1899 -----DANGRRITINS-----ASGRV-----QYEGKLNQVLK--- 1927
QY 1117 DGABCEPTDKYGHISIVTQNFPTIYIGNITACHTTFADGTED--HATKFPANPTDPCQITREY 1174
D 1928 -----ETHEDGTVI---EYTYDGFGRKRTV--TKIKGSSKTVASFNIMV-----OLTKV 1973
QY 1175 H-----HTHP--DMPDN-----RLKYDKXGRVINITDNG 1203
D 1974 NDESISYDKNGNFTSDGKFTYTDADNDLTAVYKGEDKPFATYKYDEKGNRIQKTVN--G 2032
QY 1204 NTFNFTYDTLGRLONGQGSVYGYDPLNRLVSQKTDITLDELY--YRETMVLVNEVRNEMIR 1262
D 2033 KTNVTFYDG-----DSLNLVLETPDADNNVTKSYTGD-----SQGLLS 2070
QY 1263 LKRTGETTIAQORASHVLLTGTDSQOSVILTSDKONLSQEAAYSAYK--HKSTANDA--S 1318
D 2071 YTENGKKVFYHYVAHDIIAISDSTGTV-----AKYQDAMGNPTEASDEVKDN 2122
QY 1319 ILGYNERPADPVSGVTHLNGRYSYDPTLMREHT--PDSLSPPGAGINPYSCLDPIN 1376
D 2123 RYRYAGYQYDEFTGLYLLMAY--YEPNRGVFLSDPDPSDGLSDQNGYAVAGNNPVM 2180
QY 1377 RSDPSGHLSSQAWTQIGMGIAGLLTITATGMAIAAG-----GIAAIASTSTTALAFA 1432
D 2181 NVDPDGH--W-VMLVYNAGFA-----AYDGYAYYSGKMGKMAAANAASNPBGKIFKG 2231
QY 1433 LSVTSDITSIVSGALDASPKASSILGWVSMGN---GAAGLAEASIKGCTKLATHIAGP 1488
D 2232 ASRAYFTK-----KAVKITGHTRHGLNQSIGRNG-----GRGVNLRALKNA- 2273
QY 1489 AEDGENALAKSTSESSRIKGVTRSLDR--LYRNEGOYIKDHSRGYDNPFGKEQOI 1546
D 2274 -----VRSPPKVVIKOPNGATKYVGKATVVLNRKGVITAYG-----SS 2312
QY 1547 LVHGDQDGLYHTEGNK 1563
D 2313 RAKSGHVFHTHKGK 2329

RESULT 10
ABU18641
ID ABU18641 standard; protein; 2234 AA.
XX AC ABU18641;
XX DT 19-JUN-2003 (first entry)
XX DE Protein encoded by Prokaryotic essential gene #4168.
XX KW Anticense; prokaryotic essential gene; cell proliferation; drug design.
XX OS Bacillus anthracis.
XX PN W0200277183-A2.
XX PD 03-OCT-2002.
XX PF 21-MAR-2002; 2002MO-US009107.
XX PR 21-MAR-2001; 2001US-00815242.
XX PR 06-SEP-2001; 2001US-00948993.
XX PR 25-OCT-2001; 2001US-0342923P.
XX PR 08-FEB-2002; 2002US-00072851.
XX PR 06-MAR-2002; 2002US-0362699P.
XX PA (ELIT-) ELITRA PHARM INC.
XX X Wang L, Zamudio C, Malone C, Hasebeck R, Ohlsen KI, Zysek JD,
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH,
XX DR WPI, 2003-029926/02.

```

DR N-PSDB; ACA22511.  
XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 25; SEQ ID NO 46565; 1766bp; English.  
XX  
CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 623 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC the gene product; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC identified for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 2234 AA;  
Query Match 4.3%; Score 386; DB 6; Length 2234;  
Beet Local Similarity 18.7%; Pred. No. 8.3e-17;  
Matches 339; Conservative 236; Mismatches 633; Indels 608; Gaps 76;  
QY 21 PTOANNFNSVNSGVDR-----TGLYNIOITLGHVY-----NGUL----- 57  
DB 748 YSKVGHASEEISHEFPFRPSODSGFPGMFVWASIPVNGKVNATNGNFIMSEKDITLSG 807  
QY 58 -GPTLPLTISYSPLNKTDIGFIGNFGLSV---YDRKNSLLSLG-----TGEN- 102  
DB 808 RGPVSVERTNSQSKVGLFSTGWSGLEERVMADGKNLLITDGNATITFTTGDK 867  
QY 103 -----YKVIETDKTVKLOQKQDLNLFKDKLKENCYRIHKSGLIEVL 145  
DB 868 YQAPGTIYLEIKOVGGEYIKDKQTV-----TFKSGDAQGR 905  
QY 146 TGFNNNAADLKP-----KLLNPAHAIYIDMFEATQPLNLIYDDLDHDIPL 196  
DB 906 IEYTDKRGNTTYYEYDASRLSKVKNASGKELV--GYDGNKKKAAYV-----GPDKXT 959  
QY 197 INLEYGGLIKITLTFPGQ--KEGYRETELPLNROLNSIHNSPLGNEPLTWSFGY---- 250  
DB 960 ITFNVTDGLVLSSTPEGVKRYCYD-----NGULTSIYDPOHTDAPRYKTSVAYEBDR 1013  
QY 251 -----TPIGKNGIL-----GQWITSMAPGLKEITYVNSNNQGHFPOSANLPLVYVTL 301  
DB 1014 LVKATVDPGKATVTLAVNTSGKEVTLTPNGKRYVTVYND-----AGRPV----- 1057  
QY 302 MKQVGAQGPALQAEYSYSHNYVG-----GSGNGIWNKKDLNLYGLMTEYN- 348  
DB 1058 -KTVEDVGLNLITTSYEYVANNLVKTTTPKQNETATYDNGNVTISVTDGMEKEFEYNK 1116

QY 349 -YGSRESRRYKXK-----GHDQIVRIERTYNNHLLTSECKQONGYIQTETAYAI 400  
DB 1117 DNGIKATDNEDKRTIYAVYAGNTEVSQTQGAN-----TSSVIHHQYGNPIETSELS 1171  
QY 401 IGHNFDSPQFOLPKRTETWRSADNSYRSELTETTFDSGNPLTVIYDKTKQTKIISF 460  
DB 1172 AGGNLIONPS-FEM--NGTEKVVYKVDTNNSGSIS-----KDATPAPGLGGSLSKITYTK 1223  
QY 461 STH--MEYYPAGVNDCEPEYGFTRFYVKLI-----QRPYDSFPDDP 503  
DB 1224 ATNNDWGIYIAIQETLEPMTTYTLTSGMXTDLVNGAIFRVVSGSLNENAGCIDGMDTR 1283  
QY 504 EKPIQYVYSLIGSQSHVTLKIEERHSATOLNLSLTFQYNTDKSELRL-LKQTECKGE 562  
DB 1284 HNKVQGSQDWNNRQ--VTFPTTEQ---TRKVKIYLVENGSGATSSGAWPDKIQLEKGE 1337  
QY 563 NGKTVSVHKFTYTKQ--DITLQO-SHSITTH-----DNFTIHSQVRSRYTGLIFS 611  
DB 1338 VSSSFNFVLSNSSFENNPDGFVPQWVWSCQHCERNVDSDSPFGHSSIVMER-----S 1391  
QY 612 DTDKDIVTQMSYKJGRLILTRLNSGTPYANTLTVDYELNLQDDNRPPVITTTDVNG 671  
DB 1392 EYGFNDI-----GYRNRVILNQKKAETVTLTAMSKSENVND----- 1428  
QY 672 NQLENEFDGAGRHYSOCLKDSGDKFEYTIHQYDEQGRHHTSYGSDYLTNGRQOTDP- 730  
DB 1429 -----APDKLSKDYAVLAERTYDQ-----TVVNNYISPPSGNTDM 1464  
QY 731 -----DKVHLSMSKSYDNNGQIANTWMSGVSEKITVDPITLTATQLOGNS 777  
DB 1465 NRSAAVIPAKKPLOKIKIFILFRKNNG-----KWFDPDIRLBNALKNK 1511  
QY 778 NNVTGKEVNTYTPSQPQIOTLFDGAGHLOSCHTLTRDGDWRVKEKETAIGQCTIYOYD 837  
DB 1512 YD-NDGNVAVTY-----DEBQOK--NTFTYDASGAKKSETDEGNTKLDYN 1555  
QY 838 NYNKVIQITLPDGTIVNRKAPFSTDTLITDIRVNGISLGQOTPDGSLRLTQSDGGRW 897  
DB 1556 KDNLLTKVTLKNGTSVAVRYD----- 1576  
QY 898 AYTSAGNDQCPSTVITPDGQFIHQYQPELDAVLOVASNETQOQSPSYNPYATLAKAV 957  
DB 1577 ---HNATTEKSYWFGKQTHREYVDVNDKNTVYDIALNRREN--TYDENANKIKITKM 1631  
QY 958 AEGSLTPYIPSGRLKMNENINDMKMSYLTLAGLENGYDTLGTIOKISPDTHGRVYO 1017  
DB 1632 PNGSILSVYDTADRVGK-----RNGKDSFT-----FERDONGVYTK 1670  
QY 1018 IKD--SSIKTTLVNDNLRHIGSQVTDLATGMLTTVFPDGLNREIGRKLCDSSGHTLD 1075  
DB 1671 VKDLVNGVERTKYDKADR-----VTS-AT-----DSRGKID 1702  
QY 1076 IQQSWL-----KTQOLANRIYKLVNGVLQRTBOGYSDSRRLNQ--YKCPGACPTDKY 1126  
DB 1703 ---WAHYDKANSKTEKLEQVTVQGGY---TNKVSXD-YVTLIDONIRVTVGSGQ----- 1748  
QY 1127 GHSIVTQNFYDIGNITACHTTPADGTEDHATPKF--ANP-----TDECOLTEVANT 1177  
DB 1749 ---TRFYDDQGNV---RTTYAANGSGSTNYDQANKIKOLVGVTSISILSEYE 1799  
QY 1178 HPDMPDNIRLKYDAG-----RVINITDNHGTENFTYDTLGR-----LQ 1217  
DB 1800 YDQGNRTKIKHGBAGGKVTETNFVYDPIQLNLBEVLPNGTTKSYTDGFGNRTSVKVI 1859  
QY 1218 NGQGS---VGYDPLNRLVSGKTDITDCEL----- 1244  
DB 1860 NGKETSIAATFNGNQLVKGNESLTYDVANGNRTSDQKRYTNWEDDQIVAITKQGENN 1919  
QY 1245 ---YRETMNLVNEVNGEMIRLRTGETI-----IAQ 1274  
DB 1920 AFATYKTDDEDNRRIEKVNVGOVTRYFYDGDGINPLYETDNGVTLRQYVYSGADGARLMMK 1979

QY 1275 RASKVLLTGTDSQOVI--LTSRKONTSEAYSAVK-----HKSTANDASILGNGERA 1327  
D 1980 AAGGTLYHNPRDVAAMTRQDEKVAITYEDMGNVLSDTGIAAD--NPEYIAGIMY 2038  
QY 1328 DPVSGVTHLNGVYKSYDPTLMRFPTD-----SLSPFGAG-----GINPYSYCLGDDINR 1377  
D 2039 DKEIGMTYLL-----IARYNPEHGVFLSYDPDGEDDPTVMNGTYADNNPVM 2088  
QY 1378 SDPSGHSKQAMTIGMGAGLTLITATGMAIAAAGIAAATSTSTTALAFALSVTS 1437  
D 2089 TDPGKAM-----LVPVYIAGAMVAFGAKYAI-----RYGAKYQKAKVKS 2131  
QY 1438 DITSVGALEDAEPKASILGVSMGGAAGLAESAIGKGTAKATIHGAFAEDGENALL 1497  
D 2132 -----GWDY-----GKVAKSGMNGKSIQAQIPIRIHKVR---I 2163  
QY 1498 KSTSESRIRKMGVTRSLDREIVRNEBGOVTKDHSRGVTDNFMGKEQALIVHGDKDFLY 1557  
D 2164 KGDNDKKGKGVYVISTTK-----KTGRKTS-----SFEHTPHNGHY 2203  
QY 1558 HTEGNKNG-KGPYTR 1572  
D 2204 HLOKNKYSKYQKKNR 2219  
RESULT 11  
ADCO1365  
ID ADCO1365 standard; protein; 1400 AA.  
XX  
AC ADCO1365;  
XX  
DT 04-DEC-2003 (first entry)  
XX  
DE Enterohaemorrhagic E. coli O157:H7-specific protein SEQ ID NO: 1410.  
XX  
KM enterohaemorrhagic; anti-bacterial.  
XX  
OS Escherichia coli; O157:H7.  
XX  
FN JP2002355074-A.  
XX  
PD 10-DEC-2002.  
XX  
PF 24-JAN-2002; 2002JP-00015959.  
XX  
PR 24-JAN-2001; 2001JP-00112010.  
XX  
PA (UYTS-) UNIV TSUKUBA.  
XX  
DR WPI; 2003-451640/43.  
XX  
PT Enterohaemorrhagic Escherichia coli O157:H7-specific nucleic acid molecule  
PT and a polypeptide and its use; a polypeptide, a vector and a host cell.  
XX  
PS Claim 3; SEQ ID NO 1410; 2067bp; Japanese.  
XX  
CC The invention relates to a novel enterohaemorrhagic Escherichia coli  
CC O157:H7-specific nucleic acid molecule. A polynucleotide of the invention  
CC has anti-bacterial activity. The polypeptide can be used in detection  
CC and/or treatment of O157:H7 infection. The nucleotide sequence of the  
CC genome of Enterohaemorrhagic E. coli O157:H7 was determined. The present  
CC sequence represents an E. coli O157:H7-specific polypeptide of the  
CC invention.  
XX  
SQ Sequence 1400 AA;  
Query Match 4.1%; Score 365; DB 7; Length 1400;  
Best Local Similarity 20.9%; Pred. No. 1.2e-15;  
Matches 284; Conservative 154; Mismatches 467; Indels 456; Gaps 65;  
QY 418 KTEEM-BSADNSYRSSEITETTFDESGNPLTKVI-----XDKTKOKLIISPSTMEY 466  
D 218 RTLYREDAAGDLAGEITGV-T-DGAGREFRVLVLTQAGRAEAKKQHTASLSPDT----- 272

QY 467 YPPAGEVDNCPBPYGFTRPVKKIIOTPYDSEKDDPEKFOYRYSILISQSHVTLKIEE 526  
D 273 -----PPLSDSAPFDTLPGTEY-----GDKRIR----- 297  
QY 527 RHYSATOLLNSTFOYNTDKSELGRLLKQTECKGENGTASY-----VHKFTYKOD 579  
D 298 --LSAVWLTHDPAPPSLPAPLARY-----TYTAGELLAAYDASNQVRAFTYDAQH 349  
QY 580 DTLAQSHSITTHDNFTIHRSQVRSRYTGLFSDTDKDIYQMSYDKLGRLLTRTLNSGT 639  
D 350 PGRMVAH-----RYAGR-----PEMRRYRYPDGTGVVREQNLNAGL 383  
QY 640 PYANTLTYDELANLQDDNRPPVITTTD--VNGNQLNBERDGAHRYSQGLKQSDGDKF 698  
D 384 SY--RYOYE-----QDR-----LYTDSLNRBEVLTHEGAG--LKRVKKELDAG-- 425  
QY 699 YTIHQOYDEQGHHTSTYSDYLTNGR--QOTDPDKVHLSMSKSYDNMQIANTHMSYGV 756  
D 426 -----SVTHSGYDAGRLLTQTD-----AAGRRTYGL 453  
QY 757 SEKITVDPTLTATKOLQSNNSNVQKEVTTYTPSQPIQITLFDAGHLQ---CHT 812  
D 454 N--VVSGDITDITTPDGRERKFFYNDGNQLTAVVSPDGLSRBAYDEPGRVSETRCGD 511  
QY 813 LTRDGMRVKKE-----TDALGCTIYQYDNVRVILQITLPDGTIVNRKXAPSTDTLIT 867  
D 512 VIRYAYDNPHSELPAITTDATGSTRQMTMSRYQLAFDTCGQYRREYDRPQGMAYH 571  
QY 868 DIRVNGISLQOTPDGLSRLTQSDG--GRWATYASAGNDQCSVTYTPDGPPIHYQOP 926  
D 572 --REBGS--RYRBYDNRGRLTSVKDAQGHETREYNAAGL--TAVITPPGNSEYQ-- 624  
QY 927 ELDDA---VLQVANSBITQOFSYNPVTGALLKAVASQSLTPITYSGRLMENINDMK 983  
D 625 ---DAMGKAVSTQGGILTRSEMYD-LAGRITTLTNEGSRSEFTYDA----- 667  
QY 984 MSYMLTLRGLENGVTDLTGIOKISPDTHGRVQIOKSSIKTLNLYDNLRHIG----- 1037  
D 668 LDRLYVQGRFPD-----GRTQRYHYDLTGKLTQSEDEGLVTIMHYESDRLTHRYNGE 720  
QY 1038 -----SOVTLATAGMLTTVAFPDGLNREIGRLKDCDSGHTLDI-----QOS 1079  
D 721 PAKMOYDEHGMWLTEISHSEGHQVAVHYGDKGRLAGSRQVHNPEITELMQHETE 780  
QY 1080 WLKTOOLANRIYKLVNGVLOREOYSYS-----RNRL-----N 1112  
D 781 AYNQGLANRVT-P-DSLPRVEMWLTGSGYLAKMLGTPLEVFETDRILRETVRSFGNN 838  
QY 1113 QYKCDGAECEPTDKKH-----SIY-----TONFTDIYGNIT 1144  
D 839 AVELSTYTTA---GHIQSQLNSQVYDRDYDNNDGDLVIRISGPRTWGYGSATRLIE 895  
QY 1145 ACHTTFADGTEDHATFEKAPNTPPC--QLTEVHNHTPD-----MPDNIRL-----KY 1189  
D 896 SVRTIASD-----LDIRIPATPDAGNRLPD-PELHPDSTLTAMPDN-KIAEDAHVYRH 948  
QY 1190 DKARVINITD-----NHGNTENFTYDLGRL-----QNGQSV--YGYDPL- 1229  
D 949 DVGRLTEKTRDRIAGVIRTDDBRTHHYDSQHRLVFTYRIHQGEPLVSRYLXDLGR 1008  
QY 1230 -----NRVLSQKTDYDCELYR----- 1247  
D 1009 RMAKRVRRERDILGMNLSRKPEVTWYMGDGRLLTVQDTTRIQVYVPGSFTPLIRV 1068  
QY 1248 -----ETMLVNEVRNG-----EMIRLRTGETTIAQORAS----- 1277  
D 1069 ETNGEREKAKQRSIAETTLQOBSENGHVFPALVRLDLRBEERARARVSESRAML 1128  
QY 1278 -----KVLITGDSQ--QSVILTSDKONTSQEAYSAVGKHKST 1313  
D 1129 AAGCLTVEQLARQVEPEYTPARKVHFYHCDHRGLPLALISDQNTAMRGEYDEMGQLANE 1188

QY 1314 ANDASI---LGYNGERADPVSGVTHLNGYRSYDPTLMRFTPDLSLPG-AGGINDPSY 1369  
DB 1189 ENPYLHQPRLPEQGDDESGLYNNRY--YDPLQGRYITD---PIGLAGGMLNXY 1243  
QY 1370 CLGDPINRSDPSG-----HLSWQWGTGIGMIGIALLT-----IATG---GMAI1AAGGI 1416  
DB 1244 PL-NPIIRMDPLGYNLYQLLYWHDSDYGTSSIDITGSDLLSLGHAHLGVAFAKKR 1302  
QY 1417 AAALASTTALALFALSVTSIDTSIVSGLLEDASPKA--SSILGWYMGMAAGLAESA 1474  
DB 1303 GEMLSIDICITYATACGHAIGGGINAITYSYSKSLPTSGVSNBYG-VTVGGVG----- 1356  
QY 1475 IKGSTKLATHLG-AFAEDGENALLKSTSESRIRKGYTRSL 1514  
DB 1357 -----HPAYTYVVDVDPN--ESSSTESVGIAGVDASV 1386  
  
RESULT 12  
ADG30698  
ID ADG30698 standard; protein; 1510 AA.  
XX  
AC ADG30698;  
XX  
DT 26-FEB-2004 (first entry)  
XX  
DE Xanthomonas axonopodis pv citri plant pathology-related XAC3245 protein.  
XX  
KM Xanthomonas microorganism; plant; pathology; bacterial pest; Xac; Xcc;  
XX  
OS Xanthomonas axonopodis pv. citri.  
XX  
PN WO2003089647-A1.  
XX  
PD 30-OCT-2003.  
XX  
PF 22-APR-2003; 2003WO-BR000060.  
XX  
PR 22-APR-2002; 2002US-0374620P.  
XX  
PA (AMPA-) FUNDACAO AMPARO A PESQUISA DO ESTADO.  
PI Da Silva ACR, Farah SC, Quaesio RB, Reinach FDC, Ferro JA;  
PI De Oliveira JCF, De Laja ML, Setubal JC, Furian LR;  
XX  
DR WPI; 2003-865444/80.  
DR N-PSDB; ADG30697.  
PT  
PT New nucleic acid molecule from a Xanthomonas microorganism, useful in  
XX  
XX determining the presence of Xanthomonas bacteria in a sample.  
PS Claim 8; SEQ ID NO 55; 145pp; English.  
XX  
XX The invention relates to a novel isolated nucleic acid molecule from a  
CC Xanthomonas microorganism where the nucleic acid molecule is associated  
CC with pathogenicity caused by the Xanthomonas microorganism, or its  
CC variant, that causes reduced or enhanced pathogenicity. The nucleic acid  
CC of the invention may be useful in detecting the presence of Xanthomonas  
CC bacteria in a sample, as well as in plant pathology, for example, for  
CC identifying nucleic acid molecules and proteins involved in pathology  
CC caused by bacterial pests. The current sequence is that of the  
CC Xanthomonas axonopodis pv. citri (Xac) plant pathology-related XAC  
CC protein of the invention.  
XX  
SQ Sequence 1510 AA;  
  
Query Match 4.1%; Score 361; DB 7; Length 1510;  
Best Local Similarity 21.2%; Pred. No. 2.5e-15;  
Matches 304; Conservative 158; Mismatches 466; Indels 504; Gaps 75;  
  
QY 144 VITGNNNNAFDLK-----VPKLLNPAGHAITYIMNFEATQPRLARITYDDIDG 191  
DB 250 ILTRNGNFYTFKKSANGTWTPTMDVRETLSEVQDNGTLTGQVATIDDSKEQF--DIDG 307

QY 192 HDIPLNLEYGOKIKTILTLFPQCKEYRTFLRNQINSIH-----F 226  
DB 308 -----KLNGISY-----TDGEQLTLTYAQRQLQSVYDTRGRLLFAVQADRIT 350  
QY 237 SLGNENPLWTSFGYTPIGKNGILGOWITSTAPAGLKEYV---NYSNNGHHPPQS--- 290  
DB 351 QVGLPDGMVLAYGYDSQARL-----RSVTLQTAAGAASALAGYDGN---ARPDALTT 402  
QY 291 -----ANLEVLPPVTLMKOVPAAGCPAI-QAEYSY-----TSHNVYG-----GSN 330  
DB 403 HRDEQDVYASWYTDAAQVRVRSVHDPYKIDBATIAYGNGTSYSNALGNVTRGIS 462  
QY 331 GIMNNKLDMLYGLMTEYNGSTESRR-----YXCKGHDQIVRIERTYNNHLLTSECKQ 385  
DB 463 KLGQAKVTAVQGLCEPCTVGAFRSRSYDNGYDPQEDVAFAGVATDMRYNNRGLLASKIEA 522  
QY 386 QN---GYIQTETAYVYAIIGHNPDQSPQFQLPKTKETRSAD-----NSYRSE 432  
DB 523 DNSSGQKRTLQF-----DMHPS-FRVP-TQQRIDADLIVARTSWYNSRGA 570  
QY 433 ITETTPDESGNPLTKVYKDKTKOKIISPTMWEYYPAGGEVNDCCPPEYGFTRFKKIIQ 492  
DB 571 LTVSRDPDSGGP-TRV-----TQRYCEDSD-----TAAQ---NC-----601  
QY 493 TPYDSEFKDDPEKFIQRYVSLIGSQSHVTLKIEBRHYSATQLNSTLFOYNTDKSELGRL 552  
DB 602 -----SLPG-----LLASDQARFIDIDSTSYT-----YL 627  
QY 553 LKQTECT-----KGENGTYSVHKFT-YTKQDD-----TLQOSHITTHDNFTIH 597  
DB 628 ADDASCTASASTCPHRGDKWKTNNALGQTEETLAYDAGRPISIKYNGIIVT-DTYH 685  
QY 598 -----RSOYRSRYTGLFSDTDTKOIVTQMSYDKLRLTLRLTNSGTPYANTLTYDEL 651  
DB 686 PRGWLTSKVRG-----ADASSEADKTRIDWPTG--LVKQVYQDGAFTATYD-AA 737  
QY 652 NNIQDNNRPPEVITTTDVNNGQLNEPFDGAGRHSQCCKSDGDKFTYTHHQYDQGR 711  
DB 738 HRLTLD-----ITDNAGNTVHYTLDNAGNRYEDTKDAATLK--RTLSRYVNOIGQ 786  
QY 712 HHT--STYSD-----YLTNGRQOTDPKXHLMSKSYDMNGQI-----ANTHMS 753  
DB 787 LKTOATASDPDPAVDANANATKVTDALATATQSEVDPINRSHTLQDVAGIKADYKFA 846  
QY 754 YGVSEKIT--VDPITTLTATKQLOSNNSNVQ-----TGEVTTYTPSQOPIQITLFDGAGH 806  
DB 847 YDALDNLTKYTDPKGLDTTYDNGFGLVKLTSPDTG--VTSYT-----YDSAGN 894  
QY 807 LOSCHTLTRDQWDRVRKETDAIGQCTTYOYDNNRVIIQITLPDGTIVNRKXAPSTDTLI 866  
DB 895 RAT-----QTDARGNTVASYDALNRLTKVTPFS-----SLDVTY 930  
QY 867 T-DIRVNGISLGQTFPGLSLRLTOSQDGRVMAVYTSAGNDQCSYITPDGQFIHQYQ 925  
DB 931 TYDVTQACTSG-ETf-SIGLITMQDGAITYQVNRFGD-LVRKQVTSNLTALVLYKD 987  
QY 926 PEL-----DDAVLQVANSNE--ITQOFSYVPVYTGALLKAAVEGOSLPIIYPSGRL 973  
DB 988 YTVGQRLRKMTYPRGAVVDYVRNAGQTTQYGVTRPAGS--RQVLLGNA---TYPPFSPA 1042  
QY 974 KMENINDMKMSYLMWL-----RGLNGYTDLTGTIOKISRHTGRVTOIKDS 1021  
DB 1043 AGWTYNGRLLARQYDDYRPOAIQDTRPGGLDVGF-----GPDPAGNLTALTPA 1092  
QY 1022 SIKTT---LAVDNLNRHIGSVYTDLATGSHMLTTVEPDGLNRBIRGLCDSSGHTLDDIQ 1078  
DB 1093 GNTTPEIGLGYDALGRLLG--LTGVTGTVI-----DG-----1123  
QY 1079 SWLTKQQLANRIVLNGVLQRTQESYDSR-NRLNOKKCDAGCPTDKYGSIVTONFTY 1137  
DB 1124 -----YSTDATGNRL-----STYVG--TATQITTY 1146

Qy 1138 DIVGNITACHTFADGTEHDATFKANPTDPCQLTEVHTHPDWPNDIRLKDYKAGRVIN 1197  
 Db 1147 -----PNDSHRISAN-----AGVAATIDATG---N 1168  
 Qy 1198 ITDNHGTENFTYDTLGLRONGQ--GSVGYVDPLNRLVSOQTDTLDCELYRETMVNEV 1255  
 Db 1169 TTAIGTARQYVYDTSGRMTQARRAGAV-----TMNRYN 1203  
 Qy 1256 RNEMT-RLRRTGTTIAQGRASRVL-----LTGTDSSQSVIL----- 1292  
 Db 1204 GKBEQVRRFVLTNTTYTLFDBAGHMLGDYDSNGAPKQOALWLDLPGCLLANMKLYIE 1263  
 Qy 1293 -----TSD-----KONLSQRAYGAGHKSTAND--ASIG--YNGERADPV 1331  
 Db 1264 PDHLGSPRVVYIDPRDVAVMTWSLKGARFGNTANQDPDGGAAVVDMPFGGRFPAAS 1323  
 Qy 1332 GVTHLNGVRSYDPTLMRFHTPDSLSPPG--AGGINPYSCIGDPIKNSDPSG 1382  
 Db 1324 GLNQ--NYFRDYEMATGRYGQSD---PIGEGISTYSTSLSPVKYIDVLG 1370

## RESULT 13

ABU48290  
 ID ABU48290 standard; protein; 843 AA.

XX  
 AC ABU48290;

XX  
 DT 19-JUN-2003 (first entry)

XX  
 DE Protein encoded by prokaryotic essential gene #33817.

XX  
 KM Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX  
 OS Salmomella typhl.

XX  
 PN MO200277183-A2.

XX  
 PD 03-OCT-2002.

XX  
 PF 21-MAR-2002; 2002MO-US009107.

XX  
 PR 21-MAR-2001; 2001US-00815242.

XX  
 PR 06-SEP-2001; 2001US-00948993.

XX  
 PR 25-OCT-2001; 2001US-0342923P.

XX  
 PR 08-FEB-2002; 2002US-00072851.

XX  
 PR 06-MAR-2002; 2002US-0362699P.

XX  
 PA (ELIT-) ELITRA PHARM INC.

XX  
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JW;

XX  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX  
 DR WPI; 2003-029926/02.

XX  
 DR N-PSDB; ACAS2160.

XX  
 PS Claim 25; SEQ ID NO 76214; 1766pp; English.

CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway

CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pot\_sequences

XX  
 SQ Sequence 843 AA;

Query Match 4.0%; Score 353.5; DB 6; Length 843;

Best Local Similarity 22.8%; Pred. No. 3,3e-15;

Matches 223; Conservative 121; Mismatches 308; Indels 327; Gaps 48;

Qy 542 YNTDKSELGR---LKKQTECTGKNGKTYSVVHKFTYTKODDTLQOSSHITNDFTIH 597  
 Db 32 HHTDSEGYRLDWNILERSLCTVDSMGRT-----RCHWMDAQGLVATV 74  
 Qy 598 RQVRSRYTGR-----LESDTDKDIYQMSYDKLGRLLTTLNSGTYANTLTYDYE 650  
 Db 75 RDEAGQMTTFKMSDEBERLLGMDTAQGGKRYVYDRGLH-----TEHD-P 120  
 Qy 651 LNNIADNRPFP---VITTDVANGQLRNEFD-----GAGRHSQCL 689  
 Db 121 LGRVEQOMHPVHQPETEVDAGVAMRYEDRGMLQAVSDLHQRTVGYDRH-QGVV 179  
 Qy 690 K--DSDGDKFYTHIQQDEQG--RH-----HSTYSDIYLTNGRQCTDPKHLMS 738  
 Db 180 RITDARGGDKYL-----QNNEDQMLMRHTDCSGQTAMFYDERTRLEVDABS--NSTR 232  
 Qy 739 KSYDNMGQIANTHMSVSGSEKITVDDPTTLATQQLQSNNSNVQTKGKVTYTSQQPQI 798  
 Db 233 YSTDGNGHLETVMPFADGRTERYQD-----AAGRLVKTSYPG--QI 272  
 Qy 799 TLFDEAGHLQSCHTLTFDEGMDRVKRETDAGCCTIYQYDNVNRVIOITLPDGTIVNRKXA 858  
 Db 273 TRWQ-----RDGGRRVRQTDANGRTATAYDVGRLTTLTNNGESYRFRY- 319  
 Qy 859 PFSTDTLITDIRVNGISLGOQTFDGLSRLTQSD--GGRWAAYTYSAGNDQCP----- 909  
 Db 320 -----DVLDRVTEQDPPGSGRRAVYVNMALNAVAYIGGERG 356  
 Qy 910 -----STVITPDGQFIHYQVPELDAVLOV-----ANETIQOF 944  
 Db 357 GEIRHGLERDAGRLTAKTTPB--TRTERKYAAPRLRIRRRHDAEGEBEV--RF 412  
 Qy 945 SYNPVTGALLKAAVEG-----OSLPIYVPSGRLKKNENIDMKKSYLTTLRGLN 995  
 Db 413 SYDSAGNLSEETAAQVLOHRYDVQGNRTETQWDDR-----TLRLY 456  
 Qy 996 GYTDLTGTLOKISRDTGHRVTOIKDSSIKTTLNVDLNRHIGSQVTDLATGMLTTTVER 1055  
 Db 457 G-----SGHQQI-----NLGRVISEFR----- 476  
 Qy 1056 DGLNRIEGRLLDSSGHTLIDQSWLTKQOLANRIV--KINGVLQRT---EYVSYSRNR 1110  
 Db 477 DILHREVQR---SQGR-LDMRRMYDRIGLRLTKCKGKRGVVPETFPIDREVAISGQDS 531  
 Qy 1111 LNOYKDGAECPDKGHSIVTQNFYDIYGNITAC-HTTFADGTEHDATFKANPTDPC 1169  
 Db 532 LKK-----KRSRQGVTDYF-YDTTGRITACRNEAYLDSWQYDAA--ANLDDR 577

QY 1170 Q-----LLEHHHTHPMPDNINRLKDYKAGRVINITDNHNTENFTYDTL 1213  
DB 578 QGETAQAAGSVVPEFRITRYGLH-----YRDEYGVVKEKGRNG-TQHYRDAE 628  
QY 1214 GRL-----ONGOSVYG--YDPLNRLVSOQKTDTLDCLEYYRETMVNEVNGEMRLLR 1265  
DB 629 HRLLEVAVTMGTYRRYGVYDAPGRVKEHEDABEGKPNRTTFLMDGRRLAQECRLGR 688  
QY 1266 TGETTIAAQR-----ASKVLLTSTD-SQOSVILTSKONLSQEA-YSAV 1307  
DB 689 SSSLIVYSDRSGSHEPLARVRAAPGEADEVLVYHTDVNGAPBEWTGCGNIWDEAGYQW 748  
QY 1308 GK---HSTANDASILGVNBERADPVSGVTHLNGVRSYDPTLMRFHTPDSLSPFG-AGG 1363  
DB 749 GNLTHEKETRPVQONLRFQSQYLDRETLHY--NLVRFYDPDICKFTISGD---PIGLAGG 803  
QY 1364 INPYSYCLGDPINRSDPSG 1382  
DB 804 INLYQVA-PNPLSYIDPLG 821

RESULT 14  
ABU19676 ID ABU19676 standard; protein; 1515 AA.  
XX AC ABU19676;  
XX DT 19-JUN-2003 (first entry)  
XX DE Protein encoded by prokaryotic essential gene #5203.  
XX KM Antisense: prokaryotic essential gene; cell proliferation; drug design.  
XX OS Borrelia cepacia.  
XX PN W020027183-A2.  
XX PD 03-OCT-2002.  
XX PF 21-MAR-2002; 2002WO-US009107.  
XX PR 21-MAR-2001; 2001US-00815242.  
XX PR 06-SEP-2001; 2001US-00948993.  
XX PR 25-OCT-2001; 2001US-0342923P.  
XX PR 08-FEB-2002; 2002US-00072851.  
XX PR 06-MAR-2002; 2002US-0362699P.  
XX PA (ELIT-) ELITRA PHARM INC.  
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Cairt GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-029926/02.  
XX DR N-PSDB; ACA23546.  
XX PT New antisense nucleic acids, useful for identifying proteins or screening  
XX PT for homologous nucleic acids required for cellular proliferation to  
XX PT isolate candidate molecules for rational drug discovery programs.  
XX PS Claim 25; SEQ ID NO 47600; 1766pp; English.  
XX XX

The invention relates to an isolated nucleic acid comprising any one of  
the 6213 antisense sequences given in the specification where expression  
of the nucleic acid inhibits proliferation of a cell. Also included are:  
(1) a vector comprising a promoter operably linked to the nucleic acid  
encoding a polypeptide whose expression is inhibited by the antisense  
nucleic acid; (2) a host cell containing the vector; (3) an isolated  
polypeptide or its fragment whose expression is inhibited by the  
antisense nucleic acid; (4) an antibody capable of specifically binding  
the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
proliferation or the activity of a gene in an operon required for  
proliferation; (7) identifying a compound that influences the activity of

CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
CC XX

Seq Sequence 1515 AA;  
Query Match 3.9%; Score 350; DB 6; Length 1515;  
Best Local Similarity 22.1%; Pred. No. 1.5e-14;  
Matches 232; Conservative 126; Mismatches 314; Indels 380; Gaps 59;

QY 481 YGFTRPVKKIITQPPYSEFDDPEKFIQYRYSILGSGSHYTLKIEBHATQLNSTLF 540  
DB 543 FQFGQFYAYADENGMWQWEDTDQTVRYRYDAG-----RVV-----TGTRQ 586  
QY 541 QYNTDSELSRLKQTECTGKNGKTYSVVHKFTYTKQDDTLQOSH-----ITTHDNFTI 596  
DB 587 GYHT-----GRFYEAGCT-----RVIDVDEWYTAANDSGIVTAERDPLGHCTYSMBL 636  
QY 597 HRSQVR-----SRV--TGRIFS-----DTTKDQIVT-----QMSY 624  
DB 637 GRIMARIDPLGRRRTDYDERGQLTSVBSGRTVDPEYDDEQRLNARLPNGCTIKLEY 696  
QY 625 DKGLRLTRTNGSTPVAANTLYDY-----EL-----NNLQDDRPVPIYTTDV-- 669  
DB 697 DHLSRLIART-----EPGKNTTYRYPGRGELRLVQGDRETRLDYDR-----LTLIDIEL 748  
QY 670 -NGNQLNEFPDGAARHVSQCLKSDGDKFYTTHTQOYDEGGRHNTSYSDYLTNGR--- 725  
DB 749 PTGARFRKXIDALGR-----LLEETSPDG-----HVTNYD-----YADGRANPGLL 790  
QY 726 -QOTDPDKVHLSMSKSYDNWGOIANTHWSYGVESEKITV---DPTTLTATKOLQSNVQ 781  
DB 791 SAVTRPD-----GSVSRARYN---SESLFVEMIDPLGRRTQIR----- 824  
QY 782 TGKEVTTYTSQQPIQITLFDGAGH-----LQSGHTLTRDGMRYRK 823  
DB 825 -----TYGPPD--LTLASIDAGAHATREYDHATRLTKYVNALGETTYTRYDAAGRIAA 876  
QY 824 ETDIAIGQCTIYQYDNVNRVIOITLDPDGTIVNRKYAPFSTDTLLDIRVNGISLQ--QTF 881  
DB 877 EIDMGGATRYDRNAVGRLLTKLPDG-----GQWRTY 910  
QY 882 DGLSRLTQSGDGRVMAVYYSAGNDQCPSTVITPDGQFIHQYQPELDDAVLQVANSBIT 941  
DB 911 DASDRLIEIDAGDVKLVRYDASG-----RLASAVQSEHTHTVT 949  
QY 942 QQFSYNTVTGALLKAVAGGSLTFIYPSGRLKXENINDMKMGVLTWLRLENGYDTLT 1001  
DB 950 -RFAYDR-NGELIGEDGHGELRLHVYADGQRK-RMTPRRETTYA----- 993  
QY 1002 GTIOKISRDTHGRVTOIKDSIKTTLVYDDINRHIIGSOVTDL-----ATG---HMLT 1050  
DB 994 -----DVSGLATVGG-----QLTIRRDGGRRIEAGGFVAQOQYDALGRIRROLAG 1041  
QY 1051 TTVFPGINREIGKLCDSGHTLIDQOSWLTKOOLANRIYKLVNGVLTORTQOYSTDSNRR 1110

Db 1042 PAVAFDPAADPARAL-----EQLTRQV-----YHDAAGQ 1072  
 Qy 1111 LNVKDGACPR---DKGHSTV-----TONFTDIYGNITACHTPADGEDIAATP 1160  
 Db 1073 LERVND-TGADTLTYQDERGQII CAESLQPSSEFRDAWMNI-AAHGORA----- 1121  
 Qy 1161 KPAFPTDPCQTEVHNTH-----PDMPNIRLKYDKAKRANINIT-DNHG---NTENFTYDT 1212  
 Db 1122 -----PVD-----AHNRREGGLPEQGVYATKYDKADARTEI EKVYEQGVAPKTYQWTWDG 1171  
 Qy 1213 LGFL-----QNGGQSVYGYDPLNRLVSOQTDLDCELYRET-----MLVN---EV 1255  
 Db 1172 LNLVLVVTTERGVMA-YRYDAEFNRRIEQ-----QVGREGTYKFLMDGPMALAEWIEQ 1224  
 Qy 1256 RNEEMIRLRTGETI-----IAQRAKRV---LITGTSQGSVILTSKONLSQE 1302  
 Db 1225 RDG-----TTGGVVVWHIEPQSFPLPAQSTBDGLFPLIDQIGRPTVDEQGRPWKA 1278  
 Qy 1303 AYAGVCK---HKSTANDA-----SILGNGERADPVSGVTHLNGVSYDPTLMRFH 1351  
 Db 1279 AYELMGKLPVKRPAADADGATSIDTTLRFSGQMDDEGLVYNNRY---YDPDSQYL 1336  
 Qy 1352 TPDLSLPPG-AGGIMPYSYCLGDPINRSDPSG 1382  
 Db 1337 SAD---PIGLLGARQAY-VHDPGQWIDPLG 1364  
 RESULT 15  
 ABU15135  
 ID ABU15135 standard; protein; 1397 AA.  
 AC ABU15135;  
 XX  
 DT 19-JUN-2003 (first entry)  
 XX  
 DE Protein encoded by prokaryotic essential gene #662.  
 XX  
 KM Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 XX  
 OS Escherichia coli.  
 PN WO200277183-A2.  
 PD 03-OCT-2002.  
 XX  
 PF 21-MAR-2002; 2002MO-US009107.  
 XX  
 PR 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX  
 PA (ELITR-) ELITRA PHARM INC.  
 PI Wang L, Zamudio C, Malone C, Haeelbeck R, Ohlsen KU, Zykkind JW;  
 PI Wall D, Trewick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 DR WPI; 2003-029926/02.  
 DR N-PSDB; ACA19005.  
 XX  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation or  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 PS claim 25; SEQ ID NO 43059; 1766bp; English.  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated

CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC product's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 CC  
 XX  
 SQ Sequence 1397 AA;  
 Query Match 3.9%; Score 342; DB 6; Length 1397;  
 Best Local Similarity 21.4%; Pred. No. 4,8e-14;  
 Matches 256; Conservative 157; Mismatches 382; Indels 404; Gaps 63;  
 Qy 418 KTBTM-RSADNYSRSITETTPESGNPLTKV-----KDKTKQKIS-----PS----- 461  
 Db 218 RTQTFHREAGPESGHTGV-T-DGAGRHFELVLTQARAEARQOASIGSTEPSAPDT 276  
 Qy 462 -----THEEYYPAGEVNCPEP-----YGF-----RFVK 488  
 Db 277 LPGTYERGNDGIRLSAWMILHDPYR-----ENLPAALVRYGWTFRGELAAYVDRSNT 331  
 Qy 489 KIIQPYDSBFK-----DDEKFIQYRYSLSGSHVTLKIEBRHYSATQLNSTL 539  
 Db 332 QVRSFTYDDKRYRMAVHRHTGRPE--ICRYD--SDGVTESQLNPAGLSY----- 379  
 Qy 540 FOYNTKSELGRLKQTEC--TKGNGKTYSVVHKTTYKODDTLOQSHSITTHDNFTIH 597  
 Db 380 YOYEKDRITITSLNREVLHTQEGG-LKRVVK-----EHAQGSYV 421  
 Qy 598 RSQVRSRYGRLESDPDTQDIYQMSYDKLGRLLTTLNSGTPYANTLTGYDELANMLQDD 657  
 Db 422 QSQFDA--VGRLEAQTDAAGRITTEYSPDVATGLTIT--TPDGRASAFY----- 468  
 Qy 658 NRPFFVITTTDVNGQLRNEFPDAGRHVSQCLKDS-----GDKEFYTHITQYDEQGRH 712  
 Db 469 NHHSQLTSATGPGGLIRREYDEWGRLLIQTAPDGIITRYRYNPHSDELPCATEDATGSR 528  
 Qy 713 HTSTYSDYLTNGEQQDTPDKHLMSKSYDNWQQLANTHMSGVSEKITVDPI--TLPAK 771  
 Db 529 KWTWMSRY--GQLSFTDCSGVYTRYDHRFQGVAVVHREGLSQRAYVDSRGOLIAVK 585  
 Qy 772 QLOSNNNVOTGKEVTTYPSQOP-IQTLFDEAGHLOSGCHTLTRDMDRVRKXTDAGC- 830  
 Db 586 DTQGHETRYE-----YNAAGDILT--VIAPDG-SRNGTOYDMWGRK 623  
 Qy 831 -CTI-----YOYDYNRVIQTLTDPGTVNRKVAFPSTDTLITDIRVNGISLQOQTF- 861  
 Db 624 ICTTQGLTRSMEXYDAAGVIRILTSNGS-----HTTFAYVDLDRLIOEGFGFGRTORYH 678  
 Qy 882 -DGLSLTOSQDGRWATYTSAGNDQCSVTITPQGGTIRHOYQPELDVAIVQVANSNEI 940  
 Db 679 HDLTGLIRSEBGLV-----THMYTD-EADRLTRTYNNGET 714  
 Qy 941 TQGFSNPVTGAL--LKAVAGGSLTPRY-YPS-GRLLKMEKI-----NDM---KQMSY 986

Db 715 AERWQIDE-RGWLTDISHISEGRVATYHYGDSKGRLASHELVVHHPQTNELLMQHETRH 773  
QY 987 LMTLRGLNGYTDLTGTIOKISRDTGRTVQIKOSSIKTTLNVDLNRHIGSQVTDLATG 1046  
Db 774 AYNAOGLAN--RCIPBLSLPAVEMWLTYG-----SGWLSGKWL 808  
QY 1047 HMLTTYEF--DGLNPEIGRKLCDSSGHTLIDIOQSMWIKTOOLANRIYKLVGLQRTQOVS 1104  
Db 809 D--TPLEVTYRDRLHRETLR-----SFGRYELTYATPAQLOSO--HLNLSLS-DRDYT 858  
QY 1105 YDSRNRINOYKCDGAECPDYKSHIYTONFTYDIYGNITACHTTFAD-----GTEDHA 1158  
Db 859 WNDNGELIR-----ISSPRQ-----TRSYSTTGRLTGCVHTTANLDIRIPYTTDPA 906  
QY 1159 TPKFANPTDPCQULTEVHTHPD-----MPDN-----IRIKYDKAGRYINITD----- 1200  
Db 907 GNRLPDB-----ELHPSDLSMWPDPNRIRADAHVLYRYDRHGRLTEKTDLIPGV 956  
QY 1201 ---NHGNTENFTYDITLGLONGGGSVYG-----YDPLNRLVSOKTDTLDCEL----- 1244  
Db 957 IRTDERTHRHYHDSQRLVHYTRTOYAEPLVESRYLYDPLGRVAKGVWRERKDLTGMM 1016  
QY 1245 -----YY-----RETMVNE-----VRNGEMIRLLR-----TGFTTIAQORA--- 1276  
Db 1017 SLSRKPOVTYGMGDRLTTIQNDRTRIQTIYQPGSFTPLIRVETATGELAKTORSLAD 1076  
QY 1277 -----SKVLLTGTDSQOSVITSDKONLSQEA----- 1303  
Db 1077 TLQSGGEDGGSVFPPLVQMLDRLSEILA--DRVSESRRLASCGLTVAQMOQSQM 1133  
QY 1304 -----YSAYGKHSTANDAS-----ILGYNG 1324  
Db 1134 DPVTPARKIHLVCHDRGLFLALISTEGTTAMVYAEYDEWGNLNBENPHOLOQLIRLPG 1193  
QY 1325 ERADPVSGVTHLGNNGYSYDPTLMRFHTPDSLSPFG--AGGINPYSYCLGDPINRSDPSG 1382  
Db 1194 QQYDEBSGLVY--NRHRYTDLQGRYITQD--PIGLKGMWNYQYPL-NPISNIDPLG 1246

Search completed: January 30, 2006, 09:46:15  
Job time : 140.372 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2006 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: January 30, 2006, 09:42:59 / Search time 37.5112 seconds

(without alignments)  
3687.335 Million cell updates/sec

Title: US-09-889-874A-23

Perfect score: 8879

Sequence: 1 VYIKFLKLRIRIMSDNNER.....PKIILGRTEKTVKPTFRP 1673

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

## Database :

Issued Patents AA:\*  
1: /cgn2\_6/ptodaca/1/1aa/5 COMB pep:\*  
2: /cgn2\_6/ptodaca/1/1aa/6 COMB pep:\*  
3: /cgn2\_6/ptodaca/1/1aa/8 COMB pep:\*  
4: /cgn2\_6/ptodaca/1/1aa/PTUS COMB pep:\*  
5: /cgn2\_6/ptodaca/1/1aa/RB COMB pep:\*  
6: /cgn2\_6/ptodaca/1/1aa/backfile1 pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1638	18.4	1584	2	US-09-251-645-6
2	339	3.8	1426	2	US-09-492-709A-340
3	331.5	3.7	1377	2	US-09-711-164-467
4	317	3.6	998	2	US-09-252-991A-28424
5	313	3.5	1627	2	US-09-328-352-6604
6	306	3.4	1586	2	US-09-543-681A-5329
7	294	3.3	1439	2	US-09-543-681A-7560
8	292	3.3	1626	2	US-09-252-991A-23805
9	285.5	3.2	1566	2	US-09-328-352-5542
10	277.5	3.1	974	2	US-09-252-991A-23640
11	271.5	3.1	1665	2	US-09-543-681A-4476
12	238.5	2.7	1043	2	US-08-851-567B-61
13	235.5	2.7	804	2	US-09-328-352-5545
14	223	2.5	1128	2	US-09-252-991A-31032
15	222.5	2.5	3290	2	US-09-328-352-5486
16	220.5	2.5	1183	1	US-08-447-031A-2
17	218.5	2.5	2123	2	US-08-968-685A-10
18	214	2.4	2504	2	US-09-328-352-5821
19	212.5	2.4	954	2	US-09-251-645-12
20	211.5	2.4	10182	2	US-09-134-001C-3159
21	207	2.3	2314	2	US-09-268-347-49
22	202.5	2.3	2277	2	US-09-543-681A-6124
23	201	2.3	1739	2	US-09-540-236-3739
24	199.5	2.3	2057	2	US-09-499-203-2
25	198.5	2.2	1092	2	US-09-147-405B-15
26	198.5	2.2	1004	2	US-09-268-347-30
27	197.5	2.2	1726	2	US-09-700-227-2

28	194.5	2.2	5024	2	US-09-710-279-2964	Sequence 2964, Ap
29	190	2.1	2385	2	US-09-543-681A-6304	Sequence 6304, Ap
30	188.5	2.1	2283	2	US-10-172-5024-4	Sequence 4, Appl
31	188.5	2.1	2736	2	US-09-252-991A-30227	Sequence 30227, A
32	185.5	2.1	1166	2	US-09-200-6508-7	Sequence 7, Appl
33	185.5	2.1	2383	2	US-09-492-709A-302	Sequence 302, App
34	184.5	2.1	2142	2	US-09-540-236-3459	Sequence 3459, Ap
35	184	2.1	3696	2	US-09-134-001C-5080	Sequence 5080, Ap
36	183	2.1	1222	2	US-09-206-942-37	Sequence 37, Appl
37	183	2.1	1222	2	US-10-193-764-37	Sequence 37, Appl
38	183	2.1	1228	2	US-09-206-942-34	Sequence 34, Appl
39	183	2.1	1228	2	US-10-193-764-34	Sequence 34, Appl
40	182.5	2.1	1565	2	US-08-851-567B-59	Sequence 59, Appl
41	182	2.0	1861	1	US-08-790-912-4	Sequence 4, Appl
42	181.5	2.0	2315	2	US-09-543-681A-5434	Sequence 5434, Ap
43	181	2.0	1228	2	US-09-463-402-2	Sequence 2, Appl
44	181	2.0	1228	2	US-09-889-572-2	Sequence 2, Appl
45	181	2.0	1228	2	US-09-117-447-2	Sequence 2, Appl

## ALIGNMENTS

```

RESULT 1
US-09-251-645-6
; Sequence 6, Application US/09251645
; Patent No. 6281413
; GENERAL INFORMATION:
; APPLICANT: Kramer, Vance C.
; APPLICANT: Morgan, Michael K.
; APPLICANT: Anderson, Arne R.
; APPLICANT: Hart, Hope
; APPLICANT: Warren, Gregory W.
; APPLICANT: Dunn, Martha
; APPLICANT: Chen, Jeng S.
; TITLE OF INVENTION: NOVEL INSECTICIDAL TOXINS FROM PHOTORHABDUS LUMINESCENS
; TITLE OF INVENTION: AND NUCLEIC ACID SEQUENCES CODING THEREFOR
; FILE REFERENCE: CC1963/A
; CURRENT APPLICATION NUMBER: US/09/251,645
; CURRENT FILING DATE: 1999-02-17
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 6
; LENGTH: 1584
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
; US-09-251-645-6

Query Match      18.4% Score 1638; DB 2; Length 1584;
Best Local Similarity 31.0%; Pred. No. 3.1e-118;
Matches 493; Conservative 233; Mismatches 682; Indels 182; Gaps 54;

19 EFFTOANNFSAAGGVDPRTGLNIQITLGHIVGN--GNLGPFLPLTSLYSPLNKTIDIG 76
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5 DISNAFNSFSYINTGVDPRTGQYSANINITLRPNVGSBGT--LSLSFSLTLTANG 62
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
77 FGIGFNGLSVDRKNSLSLSTGENYK--VIEDKTVYLQCKKLDNLEFEKOLKENCY 133
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
63 FGIGFNGLSLTLTKITLTFSTRANGBOFKCKPLPPNNNDLSFKDKLKLRYVK-LDSNTE 121
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
134 RIHKGSDIEVLGFNNNAEDLVKPKLLNPAGHAIYIDNPFATQPRLNRIYDDLDGHD 193
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
122 YVNNKGIIHILRIGSS--DIKTVALEPDEADPLVNSR----- 162
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
194 IPIPLNFEYQGLIKITLTL-FPGQKEGYRTELRFL-NROLNSINNFSLGNEPPLTWSFGYT 251
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
163 FALSEIKRYRTGYTKLTKANSNN---CISVEFPDDNNISAKIAFDVRNDYLITVTPVD 219
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
252 PICK-----NGILGWIYSMTAPGSLKETVYNSNNQGHHPQSANLPVLPY 298
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
220 ASGPIDSAEPKMTYQTLKGVF-PVISTFRPTGYVEIVSYKEN--GH---KVVDTEYIPY 273
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
299 VTLMKQVPGAGCPALDAEYSYTS-HNYVGGGSGNGI-WNNKLDNLVGLMTEYNYGSTESRR 356

```

Db	274	AAALTIOPGNCQPAVSKSYEYSVHNFAGISGKTSPESSODNLHYTGKTTYSIE---	330
Qy	357	YKDXEGHDQIVLIERIYNNYHLLTSECKONGYIQTETAYALIGHNFPDSQSOQLP-	415
Db	331	-RVLDDGQSVAVIERVFNKFKHMTKKAQONKKAITTEITYNEDLSXSFSSQFENLQOPS	389
Qy	416	KTKETWRSADNSRSEITETTPESGNPLTKV/KDKKTQCLIPSPHMEYYPAGEVDN	475
Db	390	RVLRRYTDIQNTSRREETVNIKSDWGMTLL-IETSGIOK-----EYUYYPNGEGNS	442
Qy	476	CPEPYGFTFRVVKKIIOTPYDSEFPDDEKFIORY-----SLIGSQ-----SHVTLKIE	526
Db	443	CPADPLGFSRLKSVTQKGSPPDAASVANKIHYTQKFPFTTGAAYKAYSXKISEIIN	502
Qy	527	RHVSATQOLNSTLFQYNTDKSELGRLLKQTECTGKNGKTYSVVHKFTYTKODDTLQOSH	586
Db	503	KIATFYSVNSP-----TSKSH-GSLAKITSVMANQ-----QYVTFKYEYSEBEMTTNA	551
Qy	587	SITTHDNFTIHSQVRSKTYGRLEFSDDTKDVIQMSYDKLGRLLTITLANSCTPYANTLT	646
Db	552	TVTFGDSGAMHESKNVTSIYTHROUKRVYDVNHVITDOSYDLGRITGQIIDPETAIREIKRN	611
Qy	647	YDYELANLNQDNRPFPVITTTDVGNGOLRNEFDGAGRIVSQCLMDSDC-----DGKF	698
Db	612	YVYOQPGSEDNDFW-VNIEVDSQCVRKKTHYDKMGRCSLEBQDDAMGTSIGYQCTY	670
Qy	699	YTHIQQYDEGBGRHHTSYSDYLTNGRQOTP-----DKVHLSMSKSYDNNGQIANTHMS	753
Db	671	RKVLARQYDVLAGOLSKESINDMLWN--LSAPDLVRLATPLVTTKYKVDGNGNLYSTES	728
Qy	754	YGVSEKITVDPITLTATK-----QLOSNNSNVQTKGEVTTTYSQOPIQITLDEAGHL	807
Db	729	DGRLEIEIHDPITRITITQGVKGLMANTQONNF-----EOPASIKAVYPDGTI	776
Qy	808	QSCHLTHFDGMDRYAKETDAIGQCTIYQUDVNNRYIQLTDPGTIVNRKAPFSNDTLIT	867
Db	777	YSTRYRYDGGRTYETETDABGHATQIGYDVPDRIVKTLTDPGIIIBSAVFSFHEILIS	836
Qy	868	DIVRNGISLGOOTPEGLSRLTOSODGRHVMAYTYSAGNDQCPSTVITPBGQFIHYOPE	927
Db	837	ALANNQTOGLALVYDGLGRVISDPTVGKRTIELVYRGQDK-PIGSITPSHKQNMDDIYLY	895
Qy	928	LDDAVLQVANSNETIQOFSYNPVTGALLKAVAGOSLPFI-YYPGRLKMEH-INDMKMS	985
Db	896	LGSVMSKFTTGTDOQNFYHSKXTGLLSA-SEGVSQTYMSFSPSGVLQRESFELDNKPKIS	954
Qy	966	---YIWLTLRGLENGTDLTGCTIOKSRDTHGAVNOIKDSSIKTLVNDLNRHIGSOYTD	1042
Db	955	SGEYLYTWSGLIOHKDSEFHNHVSYPAOGLVTEBDQAOYATEYUNVGALLITTYKD	1014
Qy	1043	LATGHMLTTFVEFDFIANEIRGLKCDSSGHTLIDIOQSWLKTQOLANRIVLKLVNLORTQO	1102
Db	1015	TTSLSQLVTKIEYDAFDEIRKSLSDPSIQV-ILTSYTKNNQISQRTSIDGVMMKHEH	1073
Qy	1103	YSYDSRNLNOYKCDGAECPYDKYSHSVTQNFYDIYGNITACHTTPADGTEDHATFE	1162
Db	1074	YQYDNRQSLTSYQCEGSEOSPIDHTGRVLNQCIYHNDQWGNIKRLDNYFRDKEET-VDYHF	1132
Qy	1163	ANFPDPCQLTVEHNHHPMDPNRIKLKYDKAGVINIIT-DNNGTENTFYTDLGRL---QN	1218
Db	1133	SO-ADPTQILRI-----TSDKOQIELSYANG---NLTEDEKQOT--LIYDQNNRLVQYQD	1183
Qy	1219	GQGSV---XGYPLRLRLVSQ--KTPLDCELYURETMLVNEVRNGEMIRLRLTGTETIAYQ	1273
Db	1183	RLGNLVCYQYALAKLTPQVLANGTVNRQ-HYASGKVTNIQLOGBEALITLSDKORIGH	1244
Qy	1274	QRA-----SKVLLTGTDSQOQSVILTSDKONLSQEAYSAYGKHKSTANDASILGNGERADP	1325
Db	1242	QSAKKGQSVYUYQYGDHNSVTIASQNMENMLASTPYGFRSLI---SSLPELNGAQYDVP	1298
Qy	1330	VSQVTHLNGHRSYDPTLMRFRTPPSLSPFGAGGINPVSYCLGDTIINSDBPGHLSQWAM	1386

```

Db      1299  VTGMYFLNGYRVENPYVLMRFHSPDSMSPCRGGINPYTCQGGPINRIDLNGHLSAGGI 1358
QY      1390  TGICMGIAGLLITLAT--GGMALA---AAGGIAAALASTSTYLAANGALSVTSDISIV 1443
Db      1359  LGIYLGAIIGIIWGVISVAGAGAAISAGLIAAGGALGAIASTSALATVATVIGLADSIGIA 1418
QY      1444  SGALIEDASPRKASSTILGVSMGMGAAGLAESAIKCKGTLATLHGAFAEDGENALLKST-SE 1502
Db      1419  SAAISEKDPKTSGLIMNIISAGLGVSFGISAI-----TFT-----SLVKSARSQ 1463
QY      1503  SSRIKMGVTRSLDREIVRNEEGQVYIKDHSR 1532
Db      1464  SQAVSAGVIGSVLEF-----GEVASRSSR 1468

```

```

RESULT 2
US-09-492-709A-340
; Sequence 340, Application US/09492709A
; Patent No. 6720139
; GENERAL INFORMATION:
; APPLICANT: Zyskind, Judith
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Trawick, John
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Froelich, Jamie M.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; TITLE OF INVENTION: ESCHERICHIA COLI
; FILE REFERENCE: ELITRA.001A
; CURRENT APPLICATION NUMBER: US/09/492,709A
; CURRENT FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 340
; LENGTH: 1426
; TYPE: PRT
; ORGANISM: E. Coli
US-09-492-709A-340

```

	Query Match	3.8%; Score 339; DB 2;	Length 1426;
	Best Local Similarity	20.2%; Pred. No. 9.6e-17;	
	Matches	Conservative 150; Mismatches 409; Indels 458; Gaps	60;
Qy	572 KFTYTKODDTLQOSSHITTHDNFTIHRSQVSRKYRGLFSDTDDTDKDIOTWSYDKLGRL	631	
Dd	317 RYTTLAEAGELL-----AVYDRNSTQVRAPTYDAQHGRMVARVAYGRPFRARYRDDDTGAVV	372	
Qy	632 TRLTNSGFYPAANTLTYYDELNNLODNRPPVIITTD-----	668	
Dd	373 EQLNPAG-----LSRYR---LVBQDR---ITVTDSLNREVLHTEGGAGLRVKKE	418	
Qy	669 -VNGNOLNEPFGAAGRVAHSQCCKSDSGEYTTITQGYD-----EOGRHHTSYSDYL	721	
Dd	419 LADGSVTRSGVDAAGRLLTAQ--TDAAGRTREGLNVASGDILTIDTPRGRETAKFYND-	474	
Qy	722 TNGRQQT--DPDKVHLMSKSYDNMGQILANTHMSYGVSSEKITVDPTLTATKLOLSNSN	778	
Dd	475 --GNQLTVAVSPD--GLESRREYDEPRGLV-----SE	502	
Qy	779 NVQGSKETL-FYTBQCPQIQLTFPEACHLOSCHLTRDGMDRVAKETPDALGOCTIYQD	837	
Dd	503 TSRSEIVRYRDIDAHSLEPATTTDATG--STROMTSRRGQLAFDDCSGTQRERYD	559	
Qy	838 NMYRVIQITLPDGTIVNRKYAPFSTDILTDIRVNGISLGOQTFDGLSRLTQSQDG-GRV	896	
Dd	560 RFQGM-----TAVHRE-----EGISL-YRRYDNRGRLTSVKDAQGRE	595	
Qy	897 WATYSAGNDQCPRIYITPDQFIHYQQ-----	925	
Dd	596 TRYENNAAGDL--TAVITPDGARSETQYDANGAKAVSTOGGLTRSMEDYDAAGRYSLTINE	653	

QY 926 -----PELDPAVLQVANSNEITQOFSYNPVYTGALLKAVAGOSLPIYPSGRKME 976  
 DB 654 NGSHSVSYALDRLVQGGPDGRTQRYHD-LTGKLTQSEDEGLVLMYDESDRTTHR 712  
 QY 977 NINDMKMSYMLTGLKLENGYDTLGTIOKISDPTGRVTOIKDS-----IKTLLAYDDL 1032  
 DB 713 TVNGEPAEQOY-----DGHGMLTDISHLSGHRVAHYVGDOK 751  
 QY 1033 NRHIG--SOYTDLATGML----- 1049  
 DB 752 GRLTGEQCYENPEPTGLMOMHETKAYNEGLANRVTPDLSPEVEMYGSGYAGMKL 811  
 QY 1050 ---TTTVEF--DGLNREIGRKLCDSGGH--TLIDIQSWLTKTOQLANR----- 1089  
 DB 812 GGPFLVETDRDLHRETVRSGSGASNAAYELSTTTPAGOLQSLNSLVYDRDYGS 871  
 QY 1090 ---YKLVNLTQTEQSYDSRNRLNQKCDGAC-----PTDKYHSHI----- 1130  
 DB 872 DNGDLVVISGPRQ-TREYYSATGRLESVRLAPDLDIRIPYATDPAGNRLPDELPDS 930  
 QY 1131 -----VTQN-----FTYDIYGNITACHTTPADG---TEDHATFPFANPTDPCQULEV 1174  
 DB 931 TLTVPMNRILAEADHYVYRDEYGRLEKTDRLPAGVIRTDDEBT----- 975  
 QY 1175 HHTHPMDNIRLKYDKAGRVINITD--NHGN--TENFTYDTLGR----- 1215  
 DB 976 HHTH-----YDSOHLVFTTRIQHGEPLVESYLYDPLGRBAKXVWRERDLTG 1025  
 QY 1216 ---LONGGSSVYGYDPLNRLVSOQTDLCELYR----- 1247  
 DB 1026 WMSLSRKEVYWGMDG--DLTLVQTDTRIQYVEGSGFTPLIRVETENGEREKQRRS 1084  
 QY 1248 --ETMLVNEVRNG-----EMIRLL-RTGETTIA-----QQRASKY 1279  
 DB 1085 LAPTLOQEGSENGGVVPAELVRLDLREELRADVSESRAMLAQCGLYEQLARQV 1144  
 QY 1280 LLTGDSQOQS-----VILTSKONLSQEA--YSAYGKHSTANDASILGYN----- 1323  
 DB 1145 EPEYTPARKAHLHCHRGRLPALLISBDGTWASAEIDENQOLNEBNPHV--YQYRL 1202  
 QY 1324 --GERADVSGVTHLNGYRSYDPTLMFHTPDSLSPPG--AGGINPYSCLEDPINRSDPS 1381  
 DB 1203 PGQOHDESGLYY--NHRHYDDPLQGRYITOD--PGLKGWMLYQYPL-NPLQOQDPM 1256  
 QY 1382 GHLSWQWMTGIGIACLLTLTATGMAIAAGIAAIASTSTTALAPALSTSDITS 1441  
 DB 1257 GLV--QTMDDARSG-----ACTGV--CGVLSRIIGSKPDSYADALD----- 1296  
 QY 1442 IVSGALEDASPKASSILGWYMGWGAAGLAESAIKGGTKLATHLGAFAPDGENALLKSTS 1501  
 DB 1297 ---ALKEQONRS-----LCNDMEYSGIYCKOTNG-----KYFASKAETDNLK--K 1336  
 QY 1502 ESSRIKMGVTRSLDREIVRNEEGOVIRKDSRGYTDNFMKGEOAILVHGDK--DGLPYHT 1559  
 DB 1337 ESYPLRKKCPGTGDRVAAYHTHG---ADSHGDYVDEFFSSDKVLVSKONNLLEAFYLAT 1393  
 QY 1560 EGKMH--NGKRP 1570  
 DB 1394 PDGRFALNNKGEY 1407

RESULT 3  
 US-09-711-164-467  
 ; Sequence 467, Application US/09711164  
 ; Patent No. 6589738  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Foreyeh, R. Allyn  
 ; APPLICANT: Ohlsen, Karl  
 ; APPLICANT: Zyeckind, Judith  
 ; TITLE OR INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERAPY  
 ; FILE REFERENCE: ELITRA.008A  
 ; CURRENT APPLICATION NUMBER: US/09/711.164  
 ; CURRENT FILING DATE: 2000-11-09

;; PRIOR APPLICATION NUMBER: US 60/164415  
 ;; PRIOR FILING DATE: 1999-11-9  
 ;; NUMBER OF SEQ ID NOS: 469  
 ;; SOFTWARE: FaalSeq for Windows Version 4.0  
 ;; SEQ ID NO 467  
 ;; LENGTH: 1377  
 ;; TYPE: PRF  
 ;; ORGANISM: Escherichia coli  
 US-09-711-164-467

Query Match 3.7%; Score 331.5; DB 2; Length 1377;  
 Best Local Similarity 20.8%; Pred. No. 3.5e-16;  
 Matches 288; Conservative 170; Mismatches 450; Indels 477; Gaps 75;

QY 211 LFPQOQEGY-RTBLRPLNQLNSIHNPSLGNENPLTWSFGYITIGKGLIGOWITSTAP 269  
 DB 126 LFPGE-DGYSRSESLVLR--GGVAKLDEGRALALMO----- 160  
 QY 270 GGLKETVNVY-----NNNQ-----HFPQSANL--PVLPLYTLMKQVPGAGQ- 310  
 DB 161 -ALPEELRLSPHRYLATNSFQGPWMLGMCERVPDEADVLPALPFRVULGLVDRGRT 219  
 QY 311 PAIOAEYSYTSNHYVGGSGNGIWNKLDNLGYLMTENYGSTESRRYKKEGHDQIVRIE 370  
 DB 220 QTFHREAAQSGEIGVTGAMRH-----FRLVL-----TQOAR----- 255  
 QY 371 RTYNNHLLTSECKQONGYIQTETAYAILGNPDSQPSQFOLPKT--KTETWSADN 427  
 DB 256 ---AEEARQ-----ALSG--GTEPSAF--PDLTPYTEYGR--DN 287  
 QY 428 SYSEITETTPDSGNPLTKVIKDKTKQKIIISPTHEVYYPAGEVNDNCPBP--YGFT 484  
 DB 288 GILSAVWLTHDE-----YP-----ENLPAALVRYGWT 317  
 QY 485 -----RFYKKIIGPYDSEFK-----DPEKFIORYSLISQSHVTLKIE 525  
 DB 318 PRGELAVVYDRSKQVRSFYDDKRYGRMVAHRHTGPE--IRRYD--SDGVTBQLN 372  
 QY 526 ERHYSATQLNSTLFOYNTDKSELGRLLKQTEC--TKGEGKYYSYVHKETTYKQDDTLQ 583  
 DB 373 PAGLSYI-----YOEXKRIITDLSLREVLHFGEGG--KRVYK----- 413  
 QY 584 QSHSITTHDNFTIHSQVRSRYTGRLEPSDDTDIYVOMGYDKLGRLLTFLNSGTPYAN 643  
 DB 414 ---EHADGSYQSFDA--VGRLAQDTDAAGRTTESYPPDVVTLGITIT--TPDGR 462  
 QY 644 TLTYVELNLIQDNNRPFYITTTDVNGNQLRNEFPDAGRHVSQCLKDS-----GDGF 698  
 DB 463 ASAFYVNNHNO-----LTSATGPDGELRREYDELGRLIQETAPDGDITRYRYDNP 514  
 QY 699 YTHHQOYDEQHNHTSYSDYLTNGRQCTDPKHLMSKSYDNMQOLANTHKSIVSE 758  
 DB 515 SDLPATETDVTGSRKMTWRSY--GQLSFTQCSGVYTRYDHRFQOMRAVHREBSLQ 571  
 QY 759 KITVDPF-TLATKQLOSNNNVOTGKEVTTYTPSQOPIQITLFEAGHQ-----SC 810  
 DB 572 YRAYDSRGOLIAKYDQGHETRYE-----YNLAGDLTAVIAPGSR 612  
 QY 811 HTLRDGMVRKRETDALIGCTIYOYDNYNVIQITLPDGTIVRKXAPSTDTLTDIR 870  
 DB 613 NGTOYDAMGAVRTOG--GLTRSMEDYDAGRVIRLTSENS-----HTFRYVDLRLIQ 666  
 QY 871 VNGISLQOQTF--DGLSRLTQSDGGRVMAVYTSAGNDQCPSTVITPDGFIHYQYPEL 928  
 DB 667 ETGPDGRTQRYHNDLTKLIRSEDEGLV-----THNHYD-EA 702  
 QY 929 DDVAVLQVANSNITQOFSYNPVYTGAL--LKAVAGOSLPIY--PSGR--KXENINDM 981  
 DB 703 DLTRHVKGETAKERQYDE-RGMLDIDISHISGHVA--VHYVYDEKRLTGRGQYVNH 760  
 QY 982 KMSYLM-----TLRGLNGYDTLGTIOKISDPTGRVTOIKDSIKTLLAYDDL 1033  
 DB 761 QTEALLMOHETRRAYVAGGLAN--RCIPDLPVAVEMWTG-----SGYIAGMKLGD-- 809

QY 1034 RHIGSVTLDTATGMLTTTVEF--DGLNREIGRLCDSSGHTLIDIQOSWLTQOLANRIV 1091  
 DB 810 -----TPLYETTRDLARETLR-----SGERYELTAIYPAQOLQSG-- 846  
 QY 1092 KLNGVLQRTBOYSYDSNRRLNQYKCDGAECPTRKGSIVYQNTYDIYGNITACHTTFA 1151  
 DB 847 HLNLSLSDHYTWNNDNGELIR-----ISSPRQ-----TRSYSTGTGLTGVHTTAA 893  
 QY 1152 DGTEDHATFKRANTDPC--QLTEVHHTHP-----MPDN-----IRLKIDKAGRVIN 1197  
 DB 894 -----NLDIRIPYATDPAGNRLPD--PELHPDSTLSMWPDMRIADAHYLYRDRHGRLTE 947  
 QY 1198 ITD-----NHGNTENFTYDTLGRLONGQ-----GSYGYDPLNRLVSOXDTF 1239  
 DB 948 KTDILPBGVIRTDDETRHRYHYSQHLVHTRTQYEBPLVESHYLYDPLGRPAKAVMR 1007  
 QY 1240 LDCEL-----YY-----RETMVNE-----VRNGEMRLRLR-----TGETI 1270  
 DB 1008 RERDLTGWMSLSRKQYTWGWDGDRLTITIGNDRTRIQTIVQPSFTPLIRVETATGELA 1067  
 QY 1271 IAQORA-----SKULLTGTDSQOSVIL-----TS 1294  
 DB 1068 KTORRSLADALQOSGDBGSSVPPVLYQMLDRLSEELADRVSEBSRRMLASCGLTVE 1127  
 QY 1295 DKON-----LSQEA-----YSAYGKHKSTAND--AS 1318  
 DB 1128 OMQOMDPVYTPPAKIKHLYHCDHGLPLALISKEGTEWCAEYBEMGNLBNENPHOLOQ 1187  
 QY 1319 ILGNGRADEVSGVTHLGNGYSYDPTLMRPHTPDLSPPG--AGINPYSCLDGDPINR 1377  
 DB 1188 LIRLPGQYDEBSGLYY--NHRYYDDPLQGRYITOD--PIGLKGMNFGYQPL-NPVTN 1241  
 QY 1378 SDPSG 1382  
 DB 1242 TDP LG 1246

RESULT 4  
 US-09-252-991A-28424  
 ; Sequence 28424, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; PRIOR FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 28424  
 ; LENGTH: 998  
 ; TYPE: PRT  
 ; ORGANISM: Pseudomonas aeruginosa  
 ; US-09-252-991A-28424

Query Match 3.6%; Score 317; DB 2; Length 998;  
 Best Local Similarity 22.0%; Pred. No. 2,7e-15;  
 Matches 222; Conservative 129; Mismatches 362; Indels 296; Gaps 47;  
 QY 504 EKPIQYRSLIGSGSHVTLKIEHRYSATQLNSTLFGYNTDKSLGRLAKQTECTGKEN 563  
 DB 112 ERASYSYALG-----LIERADGPRTDVQDVTLYAYD-----SRGNL 149  
 QY 564 GKTSVVKHPTTKQDDTLQOSHTTTDNFTIHSQVRSRYTG--RLFSDPTFDIYTO 621  
 DB 150 TQVNTALQOVTRLGIDYDRKGKGSITDANGVTSSLA-----YTGVDGLASVSTAGSTTR 204  
 QY 622 MSYDGLRLTRTLNSGTPYANTLTYYDELNNLQDNRPPFVITTTDVNGQLRNEPDGA 681

DB 205 FDYDAVGQITRYVRGDG-----SWLSYEY-----DDARR--LVAIGNLGERLEYVDTK 252  
 QY 682 GRVSOCLNDSDDSGKFTYTHNQO--YDEGR-----HNHSTVS-DYLTNGROQTD 730  
 DB 253 GNFTAPRIKIDAGS-----LVROQOMAYDBELGRLLRAVAGGQTRSPAYDINDNPFVGTINP 308  
 QY 731 DKVHLSMSKSYDWMGOIANTHMSYGVSEKITYDPIITLTATKOLQOSNNVQTSKEVTTYT 790  
 DB 309 RQ--FAHSQAFDLIDLVLVQSGDPLGGKTRLAYD-----AQDNLTEKDPRGVTTYR 357  
 QY 791 PSCQPIQITLFDPAHL-----QSCHTLTRDGMDBVRKETAIGOCITYQDNYRVI 843  
 DB 358 E-----YDGLGNLRLVSPSGTTFEENHAGVAVIRRTARAGVTVTRYDALNRLV 408  
 QY 844 QITLPDGTIVNRKAPFSTDTLTLDIRVNGISIGQOTPDGLSRLTOSODGGRWAVYYS 903  
 DB 409 ERSSP-----SDPSLDVQYRYDITLADGNGQIGRLGAIEGARDS----- 446  
 QY 904 GNDCESTVITPDGQFIHYQPELDVAVLQVANSNETQOFSYNPVTGALLKVAEGOSL 963  
 DB 447 -----LVRYID--ERGNLVEQVNSIRLDQDTLDRVTRYDDA-----NQL 485  
 QY 964 TPYYPSGRLEKMNNDMKMSYLMTLRGLNGYTDLTGTIOKISRDTHGRVTOIK----- 1019  
 DB 486 LEIGYPS-----GLAIGY-----PRNAGQOVASVTLAVG 514  
 QY 1020 DSSIKYT-----LNYDDLNRHIGSOVTLDTATGMLTTTVEPDS-----LAREIGRLCDS 1069  
 DB 515 DKAPSTLVQIAVLPFGLR-----LTWGNGITLSBEYDDOYLLOKVG----- 560  
 QY 1070 SGHTLIDQOSWLTQOLANRIVKLVGLQR-----TEQYSYDSNRRLNQYK-CDGAE 1121  
 DB 561 -----PMQSDYQ-----HDANGNIQHRBSLMGTLYQYDPLDLRTEBRGVQSG-- 604  
 QY 1122 PTDKTGHISYVQNTFYDIYENIT-ACHTTPADGTEDHATFKRANPTDPCQLTEVHHTHP 1180  
 DB 605 -----RSYAYDAVGNRTQSDNPASGGTASQDYQYA----- 636  
 QY 1181 MPDNIRL-----KYDKRAGVINITDNHNGNTENFTYDTLGRLON-----GQSGVGYD 1227  
 DB 637 -PDSNRRLTAIGAQAQVYSDANG--NLTDQDA-ARKALYDAQGLQOSVLSGQOVAEIRYN 691  
 QY 1228 PL-NRLVSQKTDITDCELYRETMVNEVBNGENIRLLRTG-----ETIIAQORA 1276  
 DB 692 ALGGRIVKLRPESTTLYLPBGDLGEAHDGSGRLRAQYVLMPLSLPLATIDADYDA 751  
 QY 1277 S-----KULLTGTDSQOSVILTSK--QNLQSEAYSAYGKHXSTANDAS--ILGNGER 1326  
 DB 752 QGKVGNPFTLLYLHGDHLDTPRLATIDASGQIAWQSDAFRGELSGQSTQVNLRFPGQY 811  
 QY 1327 ADPVSGVTHLGNGYSYDPTLMRPHTPDLSPPG--AGINPYSCLDGDPINRSDPSG-- 1382  
 DB 812 YDAESGLHY--NYFRDYDPTGKRYESD--PIGLSGVNTYGVQGAFLNRIDPLGLAA 866  
 QY 1383 ---HLNQAWTGTIGMGI---AGLLTITATGMAIAAGIAAIAASTST 1425  
 DB 867 IEIDIPKSAVDWIRPGNIRLPAGRL-----GGVLLVASISGATPPADSDT 911

RESULT 5  
 US-09-328-352-6604  
 ; Sequence 6604, Application US/09328352  
 ; Patent No. 6562958  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gary L. Breton et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
 ; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: GTC99-03PA  
 ; CURRENT APPLICATION NUMBER: US/09/328,352  
 ; PRIOR FILING DATE: 1999-06-04  
 ; NUMBER OF SEQ ID NOS: 8252  
 ; SEQ ID NO 6604

LENGTH: 1627  
 TYPE: PR  
 ORGANISM: Acinetobacter baumannii  
 US-09-328-352-6604

Query Match 3.5%; Score 313; DB 2; Length 1627;  
 Best Local Similarity 20.4%; Pred. No. 1,36-14;  
 Matches 303; Conservative 177; Mismatches 506; Indels 498; Gaps 73;

QY 75 IGGIGIFGLSVYDRKNSLSLSTGENY--KVIETDKTV-----KIQ 115  
 DB 369 ITPAMGTEPFTTHVDAQGLIODISRTYVSNLQOMDAIFGARWYPTTKLSRKPKYT 428  
 QY 116 OKQLDNRFEPKDKENCYRIIHKSGDLEVLGFNNNAFDLKVPKLNPAGHAY--ID 172  
 DB 429 SKKDKHKOYLANGLEYIC-----LDGRALDLPDLK-----GOSTYDPIEQ 468  
 QY 173 WNFEPATOPRLNRI-----YDDLGDHDIPLNLBYOGLIKTILLFPQCKEGRYTELR 224  
 DB 469 YTTVTVSDQHLHLAYGDEKRYEKYGEDYRLSYIE-----RKNGFKVALR 514  
 QY 225 F-----LNROLNSIHNSLGNENPLTWSFGYTPIG-----KNGILGQMTSMAPGG 271  
 DB 515 YDHVSTDNKTILSDILFKODDNLALHLALQLPQGLVSDIWTIKNGQLDREVLAS----- 568  
 QY 272 LKETVYNNNOGHHFQSANLPVLPVITLMKQVPGAGOPALQAEYSYTSH-----NYV 325  
 DB 569 -----YDYOQO-----DLVOATNEPASYTYOYTHLLITRYDILT 604  
 QY 326 GGGSGNGIWNKNLNLVGLMTEYNYGSTESRRYKDEKGDQIVRIERTYNNVHLITSECKQ 385  
 DB 605 HRGKMLKMDGILPTSKAIEEMADNASRSLKLEWKK-----NIRKI-----TYLDVSGNS 653  
 QY 386 QNGYIQTTEYAYYAIISHNFDSPQSPQLPKTKETWRSADNSYRSIITTTDESNGPL 445  
 DB 654 TEHYRIDGTYRIYVDPNFE--ECFPRDAKNITTHIAKDGSK--TSYTYDERGNVL 707  
 QY 446 TKVIKDKKTKIISPSTHWEYPP--AGEVDNCPPEPY--GFTPFVKKIIQTPYDSE 498  
 DB 708 TTTQDDCAT-----SYFIDEKNQUTGMADAGQWFOYDSSGMLKEI----- 752  
 QY 499 FKDDPEKFIQYRSLIGSOSHVTIKIEBRYSATOLNSTLFOYNTDKS--ELGRLLK 554  
 DB 753 ---DPLK-----HETAYV--YNMGLVLTSTIDAKGSKSLKYDDQSLIS 792  
 QY 555 QTBC---TK--GENGKTVSVH--KFTYTKODDTLQO----- 584  
 DB 793 YTDCSGKETKQYDERGRVISIEBALNQKVEFYTELTLERREPIIKGLPLNAFGQLEKI 852  
 QY 585 SHSITTHDNFTIHSQVRSRYTGLFSDPTKQIVQMSYDKGRLLTR--LNSGTPYA 642  
 DB 853 KHAADGTEEHF-IHDAE-----GRLLAHVDPKONITTYEYDEAGLILSRIDALN----- 899  
 QY 643 NTLTYDE---LNLQODNRPPFVITTTVDNGNQLNEPFDGAGRYHSQLKSDGDKF 698  
 DB 900 HKLKTKMDRLGRLLRLINE-----NGASVYFPFDVASRLV-----KELDPFGKE 943  
 QY 699 YTHHTQOYDQO-GHHTS--TYSDYLTNGROQTD-PKVLHSSKSYDNMNGQIANTMSY 754  
 DB 944 TVVH--YDEKSGQLATSIIVASAYGODLKDRAAPKRIQOPI--FDSMGRLEORTAGY 997  
 QY 755 G-----VSEKITVDPV-TLTATQLOLSNNSNNVOTGKEVNTYTSOQPIQITLLEDEA 804  
 DB 998 GHYGLBLEEKQTEFPAYDMGRILLQANNAOSNLO-----WFDAA 1037  
 QY 805 GHLOSCHTLTRDQWDRVKEITDAIGQCTI--YQYDNRYVLIQTLPGTIVNRKAYAFST 862  
 DB 1038 GNLVQEH-----QQDYKINKTAVMKHQYDEINDRIKTRPRGQYID----- 1078  
 QY 863 DTLITDIRVNGISLGQOTFPGLSRLTOSQDGRWAVYTSAGNDQCSYITTPDGFITY 922  
 DB 1079 -----W-LTYSGHVQ--SLIVNQDPSVF 1100

QY 923 QYQPELDDAVLQVANSNEITQOFSYNPVTGALLKAVAGOSLTPLYPSGRLKMENTINDMK 982  
 DB 1101 -----ERDDLREKIA-----RHYANGVSQEOQYDLAGRLKQOM----- 1134  
 QY 983 KMSYLTWLRGLENGTYD-----LTGTIOKISR-----DTHGRVYQIKDSSI-KTTLNY 1029  
 DB 1135 -----LSHEHNGYQNYGRHNNALQETQLVQLYQYKTBELTALIRDRGRNLAHYK 1187  
 QY 1030 DDLANRHISQVYDLAAGHMLTTVEPDGLNREIGRKLCDSSGHTLIDIQSWLTKQOLA-- 1087  
 DB 1188 DPGRLLE-----ASSKKEKTEFSPDPASNIL-----DSYHSQKVQSHQKLDETSYG 1235  
 QY 1088 -NRIVKNGVLOK--TEQVYSRDNRLNOVKCGAGCEPTDKYGHSTYQTFDYIGNIT 1144  
 DB 1236 YNRUV--NNVAKSYLDQOYQYDAVGOILIKOTKSQGLD-----NLEMDVGRMV 1281  
 QY 1145 ACHTTPADGTEDHATFANPDPCCQTEVHHTHPMPDNRILKYDKAGRVINITDNGN 1204  
 DB 1282 KSRNS-----QYTALEY-----RYDALGRRIQKMSKHH 1309  
 QY 1205 TENFTYDITGLRONGGSVGYDPLNRLVSQKTDITLDELTYRETMVNEVRNGEM--IR 1262  
 DB 1310 T-----GOEQON--IIVGMDG-DTLAYESTBELTKHYIYKDSFVPMLOAVYLSPIE 1357  
 QY 1263 LTRTGETI-----IAQORASKVLLTGRDSQOSVILTSDKONLSOE-----AYS 1305  
 DB 1358 LHQTPMDSRPPVNIHNDPLMKTEKGEKEDDWFYHCDHGTQOEMTDHTGAILIKWAEYK 1417  
 QY 1306 AYGRKHSSTANDASI-----LGNGERADPVSGVTHLGNGYSYDPTLNRFPHTPDS 1355  
 DB 1418 AMGECKABEAKSNPFENSEIISNNIRPQGOYFDEBTGLAH--NRVRYYSYVGRFVSKD- 1474  
 QY 1356 LSPFG-AGGINPYSYCLADPINSRSPSGHLSQAMTIGCMGIG 1398  
 DB 1475 --PIGLGGNNVYVA-KNPITWIDSKGLCSTLIANRLG-GVNG 1514

RESULT 6  
 US-09-543-681A-5329  
 ; Sequence 5329, Application US/09543681A  
 ; Patent No. 6605709  
 ; GENERAL INFORMATION:  
 ; APPLICANT: GARY BRETTON  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
 ; FILE REFERENCE: 2709.1002-001  
 ; CURRENT APPLICATION NUMBER: US/09/543,681A  
 ; CURRENT FILING DATE: 2000-04-05  
 ; PRIOR APPLICATION NUMBER: US 60/128,706  
 ; PRIOR FILING DATE: 1999-04-09  
 ; NUMBER OF SEQ ID NOS: 8344  
 ; SEQ ID NO 5329  
 ; LENGTH: 1586  
 ; TYPE: PR  
 ; ORGANISM: Proteus mirabilis  
 ; FEATURE:  
 ; NAME/KEY: UNSURE  
 ; LOCATION: (1576)  
 ; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.  
 US-09-543-681A-5329

Query Match 3.4%; Score 306; DB 2; Length 1586;  
 Best Local Similarity 22.1%; Pred. No. 4,4e-14;  
 Matches 243; Conservative 129; Mismatches 385; Indels 342; Gaps 59;

QY 499 FKDDPEK-----IOYRYS-----LIGSOSHVTIKIEBRY-SATOLINS- 537  
 DB 486 FEANPHKINTLRMKSGADRRNNALHYRYANDSLVQIHDAVYLDIRHAYDETQRLQSV 545  
 QY 538 -----TLFOYNTDKSELGRLKQTBCTGKNGKTVSVHKTYYTKODDTL---QOS 585  
 DB 546 TRHGOBEKTLVYTYTDAQO--RLVQVYTNADK-----RVTRRFQWDESGIMAMHOYA 596

QY 586 HSIITTH-----DNFIHRSQVRSRYTGRLLFSDTDFKDIYVOMSYDKLGLTFETLNSGT 639  
Db 597 TGVSSHTRKQRFDAFTIEDNEPREYVHEHMKDKRCLEHTELTYDLAQLTLTVEVGE 656  
QY 640 PYANTLLTYDVELNLLQDNRPPFYITTTDVNGNQLRNEFDGAGHVSQCLKDS---DGDG 696  
Db 657 -----TFRRMNEQQOIIIEYTNALNETWPEWD-----TSRLAKKAIAPGDS 699  
QY 697 KFTYITHQYDEOGRHHTSTSDYLTNGROQTDGKHLSKSYDMWGQ-----IANT- 750  
Db 700 WGYT-----YDERG-----NLQWMTPEQ-----OSTCYD-WDKQFAFPAQTL 737  
QY 751 -----HMSYGVSEKI--TVDP--ITLLATKQLQSNMNVQTSKEYTYTPSQOPIQT 799  
Db 738 PNGAMMEVNEHEDIRVIDPLHITRLAMDQGLCGVDAKGNETHRYNARGQLIE 797  
QY 800 LFDAGHLQSGHLLTRDGMRYRKETDALIGQCTTYQDNNRYVQITLLPDGTIVNRKXAP 859  
Db 798 QRDGSGYPT--TLTYDWMGQLRSLLTNQNETTYTSEAGLLTECLPDET--ENRYDY 852  
QY 860 FSTDTL--ITDIRVNGISL-----GQ-----QTFDGLSRLTOSQDGRV 896  
Db 853 DATQQLVIGTIDAGRHILLRRNRGQVYARDPAGHMLHFHYDTFGKQALENEQ--SEQ 910  
QY 897 WATYSAGNDQCPSTVITPP-----GQFIHYQYQ-----BELDAVLQVANSNE 939  
Db 911 YRFEPYDLHR-----LTDEHDLGGQKHGYDVMGNVQTKTTPGSPIDTPI--PLSPQ 962  
QY 940 ITQOFSINPVYTGALLKAVAGQSILPIYPSGRLKMEINMKMSYIM-----TLGLE 994  
Db 963 VT--TFGYDKV-----GRLLFRENNADY-RTEYLQYPLSVTLRRVP 999  
QY 995 NGY---TDLTGTIOKISRD-----THGRVQ--IKOSIKITL--NYDILNR----- 1034  
Db 1000 MAVMEHEKRTITKARVEYQDLTFTYDKVQGLVREBASRGYQHHTYDLGITTETELPHQ 1059  
QY 1035 -----HIGS---QVTDLATGMLTTVEF--DGLNREIGRLCDSSGHTLIDQSWLKT 1083  
Db 1060 RAFELYVYGSHLQTOQRNDEQLVLAEYQDRLHRETLR-----TSG-ALDNEGYDCR 1114  
QY 1084 QQLANRYK-----LNGVLQRTQOYSYDSRNRL-----NOYK 1115  
Db 1115 GRITHQVARKOMNASQFVTPVIDR--RYRMKRNQILERSVSQGTGEVFTAGHWYHSYQ 1172  
QY 1116 CDGACBPFDKXGHSIVTONFYDIYGNITACHTTFAQDTEBDHATFKFANPTDPCQLTEVH 1175  
Db 1173 YDPLGQLTAHLG-SVQTEHFLYDAANLL-----TRPH-----TRAP 1208  
QY 1176 HTHPDMPNIRLKYDKAGRVINITDN-----HNTEN-----FT 1209  
Db 1209 HNQVQSGDKFYRYDGFGRMVSRYEKSSSGQRHYHDSHRIIAVDIDQGLGYQRAEYR 1268  
QY 1210 YDTIGRL-----QNGQSGYVIGDPLN-RLVSGKTTLLCELYRRTM 1250  
Db 1269 YDILGRIRLEKRLWKASALANTVYVYQHHEPDEYTFGWMGMRVSHSAAPIHTVYH--- 1335  
QY 1251 LVNEVRANGEMIRLRTGETTIIAQORASKVLLTGTDSQSVILTSQKMLSEBAYSAYG-- 1308  
Db 1336 AYNDOSYTPRLARIECT-DNPLNPGQAIYYTHSSLSGLPEALNBSGEIYVWGQOYSAMHLL 1384  
QY 1309 -----XHKSTANDASILLGNGERADPVSGVTHLGNYSYDPTLMKFTTPDLSLPG-AGG 1363  
Db 1385 QORQRTPTSTFNREQLRFGQGYFDKETGLHY--NTRFYAAPDLGSGFTQOD--PIGLAG 1439  
QY 1364 INPYSYCIAGDPINRSDPSG 1382  
Db 1440 INLYAYVA-PNPLTWDPWG 1457

RESULT 7  
US-09-543-681A-7560  
; Sequence 7560, Application US/09543681A  
; Patent No. 6605709

; GENERAL INFORMATION:  
; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
; TITLE OR INVENTION: DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709,1002-001  
; CURRENT APPLICATION NUMBER: US/09/543,681A  
; PRIOR FILING DATE: 2000-04-05  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 7560  
; LENGTH: 1439  
; TYPE: PRT  
; ORGANISM: Proteus mirabilis  
; US-09-543-681A-7560  
  
Query Match 3.3%; Score 294; DB 2; Length 1439;  
Best Local Similarity 20.3%; Pred. No. 3.2e-13;  
Matches 251; Conservative 141; Mismatches 386; Indels 460; Gaps 62;  
  
QY 347 YNYGTESSRYK-----DKEGHDQIVRIERTYNNYH-----LTSECKQONGYIQ 391  
Db 376 YHPPFDKYSRLSHVDVGNBQRH-----YNEHQHLHNGCGDLNTECEYOS--FQ 429  
QY 392 TTE-----TAYYALIGHNFDQSPQF-----QLPRTKET-----WR 423  
Db 430 LAEKTYSRLTAYVQVNHQIRRLCAVFNESAQLRVEQOTNHPYRQFGWTDAGVAMH 489  
QY 424 SADNSYRSELTETTPDSSGNPLTKVJKDKTKOKIISTHMEYYPRAGEVDCNCPPEYGF 483  
Db 490 SDKYLSERSEKRMALSEN--LWRVLENKTS-----GESYRLBYDINL 531  
QY 484 TRFVKKIIQTPYDSEFPDDEKFIQYRYSLLIGSOSHYLXIEBRYASATOL----- 535  
Db 532 TR-----TAY--WHDGSTFWQLNH-----DHQIHVDRGTGKTALIMDFGLPC 575  
QY 536 ---NSTLFQYNTDKSEIGRLIKOTECTKGNGKTVSVHKFTYTKODDTLQOHSITTHD 592  
Db 576 GCRNALGHTHISEWDALGRLLSITD--GNGNQF--RWQYONERERL--IT-- 619  
QY 593 NFIHRSQVRSRYTGRLLFSDTDFKDIYVOMSYDKLGLRLTNSGTPYANTLLTYDELN 652  
Db 620 -----VFWPDNTE--SRLAYDSLGLIKEL--SPLQITRYRDFK 656  
QY 653 NLQDNRPPFYITTTDVNGNQLRNEFDGAGHVSQCLKDSGDKFYTIHTQYDEGRH 712  
Db 657 TTL--RP-----TIRDAQAGSEFLMNR--QQLRHQDCSGK--QHIMCWDDSGRV 703  
QY 713 HTSTYSYDLTNGROQTDPKVHLSWSKSYDNMGQIANTHMSYGVSEKITVDPITLTATQ 772  
Db 704 VSQ-----TVALQE-----ATEYQYDEVGHARI-----ILPD--NSTVQ 736  
QY 773 LQSNMNVQCKEYTYTPSPQOPIQITLFDAGHLQSGHLLTRDGMRYRKETDALIGCT 832  
Db 737 LAMNAAGLLTHQNDNTPCQMUY-----NAFGRTVEIDKLAHI 777  
QY 833 IYQDNNRYVQITLLPDG--TIVNRKXAPSTDTLLTDIRVNGISLQOQTFDGLSRLTQS 890  
Db 778 HHYNMAGALISINMAGRLNLRD-----AEDRLVEIRPDE--TLQYTVNAVAGRLVEE 832  
QY 891 QD--GGRYVA-----YTSAGNDQCPSTVITPPGQFIHYQOYBELDAVLQVANSNETQ 942  
Db 833 AHLQDRVFTSAPRTILLDYDAAGLVKRETLTD-----RQYQWDSNMRL--VASKQPNQ 886  
QY 943 -----QPSYNPVTGALLKAVAGQSILPIYPS 970  
Db 887 KGLEMGQANQVHTFYDALGRIRIEQGTGDDIVERNYDELNLSRLTLPQSGSLMKLLYGS 946  
QY 971 GRLKMEINIMDKKMSYMLTLRGLNGYTLTGITOKISRDTHGRTVQIKDSIKTLLNYD 1030  
Db 947 GHATAIN-----HLVDSR-----SGLITEFRRD 969  
QY 1031 DLNRHIGSQVTDLATGMLTTTVEPFGDLNREIGRLCDSSGHTLD--IQOSWLKTQOLA 1087

```

Db 970 DLHREIS-----RTQBELTQYRQYDKLGRITSTPSSKDQHPPLNGITLWAKMYDDQ-- 1021
Qy 1088 NRIKANGVLQRTQYSDSNRLN-----OYKCGAECPT 1123
Db 1022 GNLCAMBDYRWGWEYLVSQRLKVAASENLAMLYPRADNLLERPOSEMALHSPT 1081
Qy 1124 -----DKYGHSLVTQNFYDIGNITACHTTFADGDEHATPKFANPTDPCQLTEVNH 1166
Db 1082 LELSPQDKL-RQPGWHYQDAYGVNVA-----RRYRN-----OSSQTY- 1120
Qy 1177 THPMPDNIRLKYDKAGRVINITDNHNTENFTYDTLGR----- 1215
Db 1121 -----AYGDNRLLV-IANNGIKAQYHDLGRIRHKTKVENRESQVAKQETHF 1168
Qy 1216 -----LONQGSVY-----GYDPLNRLVSQKDTDLDELRYETMLVN--- 1253
Db 1169 IMOGLRLLOBDINTGKHQTCYEHSYTPPLAVIVKSSGF---HYWHHCIDINAPL 1224
Qy 1254 EVRNGEMIRLLRGTETIIAQ-----RASKVLLGTGDSQSVILTSKONLSQEAISA 1306
Db 1225 EVTNAAO-----GNITWSGKTERFGFVRSSPLSPSSPER---VMASFQNLK- 1268
Qy 1307 YGKRKSTANDASILGNGERADPVSGVTHLNGYRSYDPTLMRFTDPSLSPFG-AGGIN 1365
Db 1269 -----YAGQYFDTNETGL-HF-NTFRFYDPOIGRITMD---PIGLLGIN 1308
Qy 1366 PYGICLADPPIRSPSG-----HUSWQAMTGI 1392
Db 1309 LYQYA-ENPLGWIDPMGLSSQEMVRVRHHTSVSEGLEGI 1345

```

## RESULT 8

```

US-09-252-991A-23805
; Sequence 23805, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,768
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23805
; LENGTH: 1626
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23805

```

Query Match 3.3%; Score 292; DB 2; Length 1626;

Best Local Similarity 21.1%; Pred. No. 5.7e-13; Indels 320; Gaps 46;

Matches 204; Conservative 125; Mismatches 316;

```

Qy 597 HRSQVRSRY---TGRLFSDTDITQMSYDKAGLLTTRTINSCTPY----- 641
Db 633 HTTQVLAIRCNERGQILBAGNMLYEABRYRVDABNVILRQLAGASPFMEWEGSKQA 692
Qy 642 -----ANTLTYP--YELAN-----LQDNNRPFTVITTDVNGNQLRNB 677
Db 693 RAVHMAASFQOMSRVYWNDSVTALNADGSEBVYVHDNAR--LVROYDPGGETLRH 750
Qy 678 PDAGRHVSQCLKSDGKFFYTIHQYDEOGR-----HNTSYTS----- 718
Db 751 YDEKQGVAAE--RDPGLG-----TTEBYRQDAGRLMLLAEGEPFTYTFDGFVMSVR 803
Qy 719 -----DYLTNGR-----QOTPD--KVHLSMSKSYDNMGQIANTHMSYGVSEKTIIVDPITL 767
Db 804 GEAQWKYERNAQGDITQTDPPEGNVTHY---AYDHRGCLVEITHADGSLHQLTWNPLGQ 859

```

```

Qy 768 TATQLOSNNSNNVOTGKRYVTTTPSPQPIQITLFPDENGALQSCHTLTRDGMVRKETDA 827
Db 860 LIEQLPDGS-----VRRYRYDTLGRQITRODDEGAI----- 891
Qy 828 IGQCTIYQVNNYNNRVIQITPDGTIVNRKYAPFSTDTLITDIRVNGISLGQOFPDGLSRL 887
Db 892 -----TRQWMAAGRLSITLPGG--ASRMV-----KRN--AYKVT----- 924
Qy 888 TQSDGGRVWAAVYYSAGNDQCPSTVITPDGQFIHYQYQPELDVAIVASNEITQAFSYN 947
Db 925 SECQEGRMTRYEX-LDDLHLVSRINPDGSQLARYX--ENARLLSEIENERGERYRLD 981
Qy 948 PYTGALLKAAVEQSLPTIYPYSGRLKMNINIMKMSYMLTLRG--LE-----NGTDL 1000
Db 982 -----YHGNGLSQETGPFQDARRAYRYDLKGLLEKTEYDDGSEL 1022
Qy 1001 TGTIQTISR-----DTGRVTLQIDSSIKTTLANTYDDLNRIHISQVTD 1042
Db 1023 RTTYQRDSTGRLAKTLPDGNRVDRYRDTLGRVAVVDGTWPLAVEYDLDNR----- 1074
Qy 1043 LATGHLMTTVE--FDGLNREIGRLCD-----SSGHTL---DIQGSMLKTQOLANR 1089
Db 1075 LVREHQMATILHYAYDALGQLIHCRLLPDGNRVDRYRQGTGLSAILMGSLTRHQFGSG 1134
Qy 1090 IVKLNGVLQRTQYSDSNRLNQYKCDGACFPDCKGHSIVTQNFYDIGNITACHTT 1149
Db 1135 RERORQGSLLSOYHDEQGRLLAHQVSQRO-----RHLYORQYRYDASGNLAA----- 1183
Qy 1150 FADGTEHATFKFANPTDPCOLTEVHTHPDMPDNIRLKYDKAGRVN----- 1197
Db 1184 IEDSRKQIRSFHY-DPLD--RLIGVGETPE-----SFVHDPRGNLAQGGQFAPRQMEV 1235
Qy 1198 -----IT-----DNHGN-----TENFTYDTLGR-----LONQGSVYGYD 1227
Db 1236 RGNRLITQGDHRFDYDAGHGLVBERRGYQKLVTEYSYDQCHRLIGVSLPDGRQVYRYD 1295
Qy 1228 PLARLVSQKDTDLDELRYETMLVNEVRNGEMIRLLRGTETIIAQ--QPAKVLTLGTD 1285
Db 1296 AFGRIARQYDQ-----RNTFELML--GERLLAESGDHRYRYLYLEPD 1336
Qy 1286 SQOSVILTS-----DKQNLQSE-----AYSAYGK--HKSTANDA 1317
Db 1337 SFRPLALDDEGEPQVEPCYQDHLCTPQELTRADRLCWSARIYAGVNLKLDILEVD 1396
Qy 1318 SILGNGERADPVSGVTHLNGYRSYDPTLMRFTDPSLSPFAGGINPYSCIGDPINR 1377
Db 1397 NPLRFQGYFDAETGLHY--NRHRYVNPSTGRYLTDPPIRL--AGGINSRY--VPNPYTW 1451
Qy 1378 SDPSG 1382
Db 1452 VDPLG 1456

```

## RESULT 9

```

US-09-328-352-5542
; Sequence 5542, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5542
; LENGTH: 1596
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5542

```

Query Match 3.2%; Score 285.5; DB 2; Length 1596;

Best Local Similarity 18.1%; Pred. No. 1.8e-12;  
Matches 202; Conservative 179; Mismatches 387; Indels 349; Gaps 49;

```
QY 440 ESNGLTKVTKDKTKTKQKISBETHWERYPPAGEVNDCEPPEYGTTRVVKIKIQFPYSEF 499
Db 337 QSGKSIYSIGAERVO-----HADYLP-----KIGFS-PIRQ-----YNSQM 373
QY 500 KDDPEKFIQVRYSL-----IGSOSHVTLKIEERHYSATQNLNSTLFOYVTDKSELGRL 552
Db 374 DEFDQSVGAKMMPPFSNMIOQNAQGYLFIDSKRKH---QLPVSIIFFETVEYVE-GWI 429
QY 553 LK-----QTECTKENGKTVSVHKEFTYTKODTLQOSHSTTH----- 591
Db 430 IKPLKNEILLIDPEGEMRSHQSDGGGNVYLKVMKMTSGEELILE-YILLDIHAYLKV 488
QY 592 DNFTIHSQVRSRYTGLFSDTDTKDIYVQMSYKGLRLTRTINSGTPVANTILTYEEL 651
Db 489 INFPLKQAEYELKE-----AFNEQVYKIAVPL----- 515
QY 652 NNLQDDNRPPVITTTDVNGNLNEPFGAGRHVSQCKQSDGDKFTYTHTOQYDQGR 711
Db 516 ----DDKAEPILA-----RYEYDTQ-----NLKAIQDNG-----HTRTYENQF 551
QY 712 HTSTYSADYLTNGRQOTDPDKVHLMSKSYDNMGOIANTHMSYGVSEKITVDPITLTATK 771
Db 552 HQLTRYTD--RTGQGNIIRYESTEAKAKAIEWADDSFH-----TKLKWHP-----RLR 599
QY 772 QLOQSNNNVQTGKQVYTTYSQGIQITLFPBAHLQSGHT-----LTDGMDRRRKE 824
Db 600 QVA-----VDAYVVPY-----YFPDLDFYTRTLADRESYTSBGKRTTRQ 845
QY 825 TDAIGOCITYQYDNVNRVIOITLPPGTIVNRKVAPFSFTDLITDIRVNGISLGQOTPDGL 884
Db 646 IDPFGRETQGYRYNODOLVKIVQNGGIIIRAY-----NKQ 681
QY 885 SRLTQSD-GRVWAAVYTSAGNDQCPSTVTPDGFHYHQ-----PELDDA---VLQV 935
Db 682 GNLVBIKQPEGSIMKREYDENRN--VSKEINPLGHITQYKNNNDQLEVIDAKAGVYK 739
QY 936 ASNETTQGFSPNPTGALKAVABQSLTPYIYBSGRKAMENINDMKMSGLMLTLRGLN 995
Db 740 QYNELGQMI SYTD-----CSGKSSYWEYDEBGLTAQYANAKVQYFYSTKGDK 790
QY 996 GYT-----DLGTIOKISRDTGRVTOIKDSSIKTTLAYDNLNRH----- 1035
Db 791 GOLQSIITPDGAKYEFHDEBGRLLK-HTDTKGLVTEYKKNQVGLLEBRIDANHSVAYQ 849
QY 1036 -----IGSOVTLATGH---MLTTVEPDLNREIGRKLCDSSGHTLIDIQSWL 1081
Db 850 WDKQGRIOKLINQNAQAYLFGYNPYGYLIRBOAFGEKHYs---YENGRGLFOIRRPNI 906
QY 1082 KTO-----QLANR-IYKLVGLORTHOYSYDSNKNRANQYKCGAECPTKYGHSI--- 1130
Db 907 LTOPDYADGOIASKSFTHLHGTQOTQOFDYNLSQLSRASNEVSQ--IDLRYNALQOL 964
QY 1131 -----VTQNFYDYIGNI-----TACHTTPADG-----TEDH 1157
Db 965 VREHONYKIPBLKPLTALVHAYDELGNLIKIRPDGHTLNLVYGSHTIYALGLANBEV 1024
QY 1158 ATF-----KRNPTDPCQULEVHHTHPDMPNIR 1186
Db 1025 VSPQDLDLHRETRRLANGLMQTKQYNDVGLLSQFNQPEQETQYLYQYQNR-----K 1078
QY 1187 LKYDQAGVINITDNGHTENTFTYTLGRLONGQ-----SVYGYPLNRILVYSQKIDTLD 1242
Db 1079 YHYDNKYLISQVEDSRIGKLANQYDPIGLLAAQSLHKTESFNFPAGNLLIDSESVSPA 1138
QY 1243 ELYURETLMVNEVRGEMIRLRTGETTIAQORASKVLTLTGDSQOAYILTSDKONLSQ 1302
Db 1139 QI-----KNMLIKSYKKGHYQYDQGVTEIIO--AGKMLKLTWQNDQNRILRSNNGLVLEY 1193
QY 1303 AYSAYGK--HKSTANDASILGVNGRA-----DPVSGVTHLNGVRS 1342
```

```
Db 1194 GYDVPGRRLYKKTAKELTLFGWDDMLWESFKSAQNTYTKHYIYEPDSFVPLQAGYKD 1253
QY 1343 YDPTLMRPHPTPDSLSPGAGGINPYSGCLDPI-NRS 1378
Db 1254 F---IQIETPD-----YQETQTKPYS-IYKDPVWNNR 1282
```

```
RESULT 10
US-09-252-991A-23640
; Sequence 23640; Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23640
; LENGTH: 974
; TYPE: PRG
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23640
```

Query Match 3.1%; Score 277.5; DB 2; Length 974;  
Best Local Similarity 20.4%; Pred. No. 3.2e-12;  
Matches 218; Conservative 141; Mismatches 377; Indels 335; Gaps 48;

```
QY 504 EKPIQVRYSLIGSOSHVTLKIEERHYSATQNLNSTLFOYVTDKSELGRLKQTECTKGEN 563
Db 112 ERWASYSYNALG-----LTKRADGPRIVQDVTTLAYD-----SKGNL 149
QY 564 GKTVSVHKEFTYTKODTLQOSHSTTHDNFTIHSQVRSRYTG--RLFSDTDTKDIYVQ 621
Db 150 TQVYNALGQYTRLGQDYDERKPGSITDANGVTSILA-----YTVGDMLASVSTAGSTTR 204
QY 622 MSYDKLRLRLTRINSGTPVANTITYEELANLQDDNRPPVITTTDVNGNLNEPFGA 681
Db 205 FDYDAVQOITRYVTGDS-----SWLSYEV-----DDARR--LVAINGNLGRLLDYDVTK 252
QY 682 GRHVSQCKQSDGDKFYTHITQO--YDEQGR-----HHTSTYS-DYLTNGRQOTDP 730
Db 253 GNRTAQIKDASGS-----LYRQOQWYDELGRLLRAVGAGQRTSPAYDLANDYGETNP 308
QY 731 DKVHLMSKSYDNMGOIANTHMSYGVSEKITVDPITLTATKQLOQSNNNVQTGKEVTTYT 790
Db 309 RO--FAHSQAFLDALRVLGSDPLGKTRLAYD-----AQDNLTEVQDPGVVTRY 357
QY 791 PSCQPIQITLFDKAGHL-----QSGHTLTRGMDKVRKETAIGQCTIYQYDNVNRV 843
Db 358 E-----YDGLGNLTRLVSPDGGTYTFEHDAAQVIRRTDARGAVTEYRYALNLT 408
QY 844 QITLPDGTIVNRKXAPSTDTLITDIRVNGISLGQTFDGSRLTQSDGGRVWAAVYTSYA 903
Db 409 ERRSPDSLSLVQR-----YDLRADGK-----GIGRL-GALDGR----- 444
QY 904 GNDQCPSTVITPDQGFHYHQYQPELDAVYLQVANSNETQGFSPNPTGALLKAVABQSL 963
Db 445 -----DSLVTYRD--BRGNLVEGVRSIRLDQOTLDRVTRYDYAA-----NQL 485
QY 964 TPYIYPSGRKAMENINDMKMSYMTLRGLENGYTDLTGTIOKISRDTGRHGVTOIKDSSI 1023
Db 486 LEIQYPS-----GLAIGY-----PRNAGGVASV----- 509
QY 1024 KTTIANYDNLNHSQVTLATAGMLTYYVFPQGNREIGRKLCDSSGHTLIDQOSWLTKT 1083
Db 510 TLAVDKAPSTLVQIQLVPLPFGULRLT-----WQNGITLSREYDOD 551
```



```

QY 1084 QOLANRIV-----KLVGLOR-----TEQSYSDSRNRLNOYK-CDGACPTDKY 1126
D 552 YOLLROKVGWQSDYQDANGNIOQRHSLMGLTIDYQDPLDRLTERGVGG-----604
QY 1127 GHSIVTGNFTYDIYGNIT-ACHTTFADGTEHDATFKCANPTPCQLTEVHHHTDMPDNI 1185
D 605 -----RSYANDAVGNRTQSDNPASGTSASQDYQYA-----PDSN 640
QY 1186 RL-----KYDAQGVINITDNGHNTENTYDTGLRLQN-----GQSYGYDPL-NR 1231
D 641 RLTAIGAQAATSDAG---NLTPDRA-ARKLAYAQAQLQSLSLDGOQVAEYRNLQGR 696
QY 1232 LVSOQDTLDCELTYRETMLVNEVRNGEMIRLRTG-----ETIINQGRAS-----1277
D 697 IVKLTPSVVTVYLYGPPGQILGEAEHDSGRKLAQYTLMLDSPLATIIDADYAOQKVG 756
QY 1278 --KVLITGTDSQOSVILTSK-ONLSQEAAYSAYGKHKSTANDAS--ILGNGERADPVS 1331
D 757 NPTLLYHGHDLDPRLATDASQIAWQSDAFGRGBALSGSTQVNLRFPGQYDAES 816
QY 1332 GVTHLGNGYSYD-----TLMRHTPD-----SLSPGAGGPNP--1366
D 817 GLRY--NYFRDYDEYETGRYVESDVEYLRKLNEMPTFLNSGESMLQATPYMEHGTPEH 874
QY 1367 -VSYCLGDPINRSD-----PSGHLSP-----QAWTGIQMGIAGLLTIAT 1405
D 875 NYTYSNNPFAKSKHGLSPNPTNLYTPDPTCTCLECKKTKTGKSKP-----925
QY 1406 GMAIAAAGIAMAIASTTALAFGALSVTSDITSIVSGALBEDASPRAS 1456
D 926 -----LVGALCSKATPFF-FGGVVCNSTIWMICGASCQECNRAP 965

```

```

RESULT 11
US-09-543-681A-4476
; Sequence 4476, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543, 681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 4476
; LENGTH: 1665
; TYPE: PRN
; ORGANISM: Proteus mirabilis
US-09-543-681A-4476

```

```

Query Match 3.1%; Score 271.5; DB 2; Length 1665;
Best Local Similarity 20.7%; Pred. No. 2,4e-11;
Matches 233; Conservative 148; Mismatches 350; Indels 395; Gaps 66;

```

```

QY 508 QRYSLIGSQSHVTLKIEBRYSATQILNSTLFOYNTDKSELGLLQOT-----EC 558
D 608 QHRIDPIYNEQSOLVSV-KRHIDGIALIL-----EYQDK-LVVICSESTTSRPRDVR 659
QY 559 TKGENGKTVSVMHKFTYTKDDTLQOSHSTTHNFHISQVRSRYTGR---LPSDITD 614
D 660 EYDTHG-----YLSQCHAYQONH-----LWHRYSPEGLAVAWGDTD 695
QY 615 TKDITVQMSYDKLGRLLTRTLNSG-----TPYANTLYT---DYELNLTQDDNRPP 661
D 696 STELT-IDYDEQGRVAVATSPSGFWMDFIYDYQMTTYIDAEQFSRYVNDN---750
QY 662 FVITTT-----DVNGQNLNEPFGAGRHVSQCLD---SDGDKFY 699
D 751 -LVTRITDPLMRETYTEMEQRKIAEINEIGERTY---GVHYVGLLAVIYLLPDKAIY 806

```

```

QY 700 TIHQYDEQGR--HHTSTYSDLYTNGRQOTDPDKVLMSKSKYDNMGQIANTHWSYGS 757
D 807 -----DYNDYQGLHFTSAGD-----EMQSLYDENGNTL-----836
QY 758 EKITVDPITTLATKQLOSNMNVQTKREVTT-----YTPSQQPIQITLFEAGHLQS 809
D 837 --IYTD-----QGRQVVEYSGHGLKALITPNGAQWQYHNPAHLKTT--NRYQSTE 889
QY 810 CHITLRDGMVRKEDTAIQCTIYQ-----YNNRVY 843
D 890 YHS---DELRLHYTDLMLHHTTRYQYSTERASTNGSLSKILLPDGVEQOQIDYSEBRVI 946
QY 844 QITLPDGTYNRKAPRSTPLTIDIVNGISLGQFTDGLSRLTOSQD--GGRVAATY- 901
D 947 AVTDEGSKTRRYGPF--DYLLAMIRPDSEI--REYISLITRLKQVYVANGAEVLYERD 1003
QY 902 SAGN-----DOCPSTYIT--PDGFIHYOYQPELDVAVLQVANSNE-----939
D 1004 KAGQIIEVDPTGREGICRYRDLGRRIATRYPDHMLRWY---NBSGLVEQSEWPE 1059
QY 940 -----ITQOFNPTVYGLAKRAVAGQSILPIYPSGRLKMENTINDMKMSLYLTLRG 992
D 1060 EBSRCLSTAQSYN-ARQOLIKATNPDSVVEFEYDQGRLCSEKINE--QEIYHQM-----1113
QY 993 LENGYDTLGTIOKI--SRDTH-----GRVTOIKDSIKITTLVNDLNRHIGSVTLAT 1045
D 1114 --NEADNTIALTRFGRRELIHYAFGLGELTSLD-----VNOHAPLOFS-----1154
QY 1046 GHMLTTFVRPDGRLNREGRK---LCDSGHT---LDIQSMWLTQOLANRYKLVGL 1097
D 1155 -----YNAVQELRLRSRAGFVNSHYTATGLAHQRAQGTGF-----LQSIQ 1159
QY 1098 QRTBQ-----YSYDRNRLNOYKCDGACPTDKGHSIVTONFTYDIYGNITACH 1147
D 1200 AHPQPEPCTDVHRSYQYD-----RAYNVGIE--DDRMRQT---RYHVNANDQITETQ 1248
QY 1148 TTFADGTEHDATFKCANPTPCQLTEVHHHTPD---MPDNIR-----LKYDKAGRYINI 1198
D 1249 YSPWKNQDE-KFOYDNNLA---ITE-HLTPSSSWPSSDAQMLQFPQQQAGRYTR 1303
QY 1199 TDHNGTENTFTYTLGRLQ-----NGQGSV-----YGYDP 1228
D 1304 YTKAG-YQDYHYDNGRLAKKIYHTRGFRPREWRYLNTQNLACPTPGDCWHYTYDA 1362
QY 1229 LNRVLSQKDTLDCELYRETMLVNEVRNGEMIRLRTGTTI-----AQO- 1274
D 1363 FGRRLS-KTIVTSDILAHIDPLPQIKITTRYIYMSGDLTEETPIYADGTLANAQOV 1421
QY 1275 ---RASKVLLTGTDSQOSV--ILTSDKONLSQEAYSAYG-----KH 1310
D 1422 QMLYQPELTPYIARYQGGKLYVYTDHQGTREIFSEGGQASWAGRLNTWGMQFMYRD 1481
QY 1311 KSTANDASI-----LGNGERADPVSQVTHLGNGYSYDPTLMEHRTDLSLSPG-AGGIN 1365
D 1482 GKXENDPNYTECPFRFAGYDEDESSGLY--NFRFRYDRRTGQYLSDD---PLGLGLGN 1536
QY 1366 PYST--CLGDPINRSDPSGHLSWQAWTGIQMGIAAG---LLTIAT 1405
D 1537 PYGVYHC---PTGMWDP-----FGLAGDCDDELIVLAT 1566

```

```

RESULT 12
US-08-851-567B-61
; Sequence 61, Application US/08851567B
; Patent No. 6528484
; GENERAL INFORMATION:
; APPLICANT: Ensign, Jerald C
; APPLICANT: Bowen, David J
; APPLICANT: Petell, James
; APPLICANT: Patisig, Raymond
; APPLICANT: Schoonover, Sue
; APPLICANT: ffehench-Constant, Richard
; APPLICANT: Rocheleau, Thomas A.

```



```

1 FILE REFERENCE: GTC99-03PA
2 CURRENT APPLICATION NUMBER: US/09/328,352
3 CURRENT FILING DATE: 1999-06-04
4 NUMBER OF SEO ID NOS: 8252
5 SEO ID NO: 5545
6 LENGTH: 804
7 TYPE: PAT
8 ORGANISM: Acinetobacter baumannii
9 US-09-328-352-5545

```

Query Match	2.7%	Score 235.5	DB 2	Length 804
Best Local Similarity	20.8%	Pred No. 4.4e-09		
Matches 147	Conservative 32	Mismatches 253	Indels 215	Gaps 30

Qy	699	YTHHQVDEGGRHHTSYSDLYTNGROOTDVKHLSMSXSDWNGOIANHNS	753
Db	21	FTLOTTEVDAGAIRSNTQAGRVTXKXET	71
Qy	754	---YVSEKITVDPITLTATYKOLGSSNNVNQTEKVEVTTYPSOQ	808
Db	72	LSNVLGGOHQHGLPIEV	116
Qy	809	SGHTLTBDGMDRVKRETDAGICITLOYODNNRVQIILPSTYINRKYAPSTDTLT	868
Db	117	TEHTIYVDPFGNITQHTADGVIYSKQYDDBGR	171
Qy	869	IRVNG	890
Db	172	RVTYGGQLSRIEKYNGDGLLISSEDKLSNKSIIINSKXDAFAGNLI	231
Qy	891	-----QDGRVMAITYSAGNDQCBSTYITD	921
Db	232	TSSYVDFPRPITVNDNGSVVTTYCYGSCGAKGAIQYDTSFGTSESNLLAAGDFADIKT	291
Qy	922	-----YQYQPELDDAVLQ	956
Db	292	LVARKGTDGVSFQUTTEFENMLKRVAVSGSSTQSYTNNSTLATEKONISIQKTFX	351
Qy	957	VAEQSLTPYYPGRLMKNINDMKMSYLTMLGLENGT	1009
Db	352	YDDGRITSIHTHPPSSVETIKYFQKDLIASRTWEVEVTVSYSLAGLAKTYYNANISEA	411
Qy	1010	---PTHGV	1044
Db	412	FDDLTGIVISHTKQKINADTNNSVYVARGNOL	466
Qy	1045	TGHMLTTVEBDGLNREIGRLCDSGHTLDIQOSWLTKOOLANIVKLANGLOR	1101
Db	467	LGEVTSJPNVIGSLINYNARQOULTVQANTDILMSYTDNSGLNNISAMTS	524
Qy	1102	QYSVDSRRLNOYKCDGACEPTDKGHSIVONFYDLYGNITACHTFADGTE	1166
Db	525	DYGDVKLNRVKL	567
Qy	1161	KFANPTDPCQLEVHTHPDMDPNIRLK	1210
Db	568	QYSVNND	616
Qy	1211	DTLGRLONGGSSVGYPIPLANLVSQKDTLDLCELYRETMVLYEVRN	1257
Db	617	DAMENAVNTDGVRTLTYDMSRLNKNQNTY	660

RESULT 14  
US-09-252-991A-31032  
Sequence 31032, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252.991A

```

: CURRENT FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIOR FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/094,190
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ. ID NOS: 33142
: SEQ. ID NO 31032
: LENGTH: 1128
: TYPE: PRT
: ORGANISM: Pseudomonas aeruginosa
: US-09-252-991A-31032

```

Query Match 2.5%; Score 23; DB 2; Length 1128;  
 Best Local Similarity 19.6%; Pred. No. 7.5e-08;  
 Matches 258; Conservative 151; Mismatches 467; Indels 442; Gaps 68

Qy	235	NE\$IGNRLPLTWSGATP-----IKNGILGOMITSMAPGLKXEVNYSNNQGH--FP	288
Db	63	D\$ALAAPPTTFAAGYSSNPRIBLQ-RGMWMLGQESMHELSEBACYLVAQGGRIQF	121
Qy	289	Q\$ANL\$PVL\$PVTLMKQVPGACOPALQAE\$YSTSHNVYGGSGNGIWNMLDNL\$GLMTEYN	348
Db	122	ALA-----PGA-----QHS-----GSEBEL-----	137
Qy	349	YGSIESRRYKQKEHQIVRIERTYNNHLLTSECKQONGYIQTETAYVALIGNF\$Q	408
Db	138	-----LRRGSSG\$BAQMR-----GRWA\$PAB-----LQTOEGSVLVL\$GHSY---	177
Qy	409	PSQFOLPKTKIETRSADNSYRSELT-----ETTFDE\$GNPLT\$KIDK-----TQ	455
Db	178	-LHQ-----RCPDGIWRLQAS\$Q\$RAGYRIEFKRS\$GGLTGV\$D\$SAG\$R\$ALVYQ	227
Qy	456	KIIS\$P\$THWEYYPAGEV\$NCPPEYGFTRFVKXIIQTPYDSEKD--DPEK-----EQ	508
Db	228	QAC\$PS-----EGDD-----GLRLE--GVIIASHD\$PPDYIDPQ\$PGLMDLVR	269
Qy	509	YR\$VLLISQ\$HVLTKIEBH\$SATOL-----NSTL\$QVNTD--KSEL\$GLLK	554
Db	270	YQ\$FSD\$G---L\$AVBDRLQOVVR\$PAMREHMLVANGEPG\$LEVR\$EMD\$HAPGR\$VYK	325
Qy	555	QTECT\$KENGKTY\$V\$V\$KFTYTKODDTLQO\$SH\$ITTDNFTIHR\$QVRSR\$Y\$GL\$F\$SDT	614
Db	326	QIEAG-----GLTRFRLR--DATEVSD\$IGRBERYFPA\$GCGRRATLVR\$AD\$S	375
Qy	615	TKD\$IVTQ\$M\$DK\$GRLLT\$RLT\$N\$GTPYANTLTYBELANLQDNRPPFVITTTDYNQOL	674
Db	376	R-----SEFDDYDL\$F\$GRL-----VAMRD\$PL\$KET	398
Qy	675	RNEFDGAGRH\$Q\$CLQ\$DDG\$K\$EYTIHQOYD\$Q\$RHHT\$S\$Y\$DYLNGRQO\$D\$PDK--	732
Db	399	RRR\$D\$GGRMLDE--\$S\$PGARY-----RKRVD\$E\$TGL	429
Qy	733	-VHL\$MSK\$YDNMQIANTH\$W\$V\$V\$SEK\$ITVDPI\$TLATKOLQ\$NSNNVQ--TG\$KEV\$TY	789
Db	430	LVELB-----DAMQ\$RMTFERDER-----GNATLVR\$PAG\$TRAY	465
Qy	790	TP\$SQ\$PIQIT-L\$P\$EAGHLOS\$CHTL\$TD\$QMDR--VR\$K\$D\$AIQ\$C\$ITTYQYDNR\$R\$VQI	845
Db	466	EDPRL\$PR\$PRI\$VDPGR\$R-----RLEWNR\$F\$GLA\$L\$T\$C\$Q\$W\$R\$YD\$D\$N\$E\$GR\$LVAS	519
Qy	846	TL\$P\$GT\$VNRKYA\$P\$STDLTITD\$IRNVG\$ISL\$Q\$OT\$P\$Q\$LSL\$T\$Q\$OD-----G	894
Db	520	SDP\$G\$Q\$LTRR\$YD\$PLG--Q\$IGLELAD\$G\$AL\$S\$Y\$E-YD\$LG\$Q\$T\$R\$AD\$E\$G\$H\$ATL\$F\$SW\$HG	576
Qy	895	RWA\$Y\$T\$SAG-----NDQ\$C\$P\$T\$VITP\$D\$Q\$F\$IH\$YQ\$P\$E-LDDA\$V\$LQ\$V\$A\$NE\$ITQ\$P\$SY	946
Db	577	DL\$AR\$V\$DAG\$G\$EL\$SYL\$H\$E\$AGRL\$VAL\$TN\$G\$VQ\$A\$F\$R\$YD\$LDL\$V\$E\$T\$G\$F\$GR\$R\$Q\$RY	636
Qy	947	NP\$V\$G\$LLK\$V\$A\$E\$Q\$SL\$PI\$Y\$P\$E\$GR\$K\$MEN\$IN\$M\$K\$K-----SYLW\$TL\$RG-----	992
Db	637	NA\$DE\$L\$AR\$E\$D\$Q\$RE\$T\$TY\$V\$DR\$G\$RL\$AS\$IR\$V\$P\$E\$TH\$AP\$L\$V\$E\$R\$Y\$RL\$AD\$RL\$AS\$AG\$AD	696
Qy	993	LEN\$GYT-DLT\$G\$TIQ\$K\$SRD\$H--GRV\$Q\$IKD\$S-----SIK\$T\$LY\$NDL\$NRH\$IG\$Q\$V\$TDL--	1043

```
Db      697 CEPRITTDVGNL-RLESGVHADQVYVSEHSHDALGVRQTSRYGD-----APPVAMLTLY 750
Qy      1044 ATGHM--LTTVEF---DGLNREIGRKLCDSSGHTLIDIQSMKLTQ-----LAN 1088
Db      751 GPGHLGALVGAVALAFERDALHREVRDA-----RRDQDALLFTQERQHAFLGRQOR 804
Qy      1089 RYKLNVLQRTQYQVSDSRNL-----NOYKCDGACPTDKYGHSL-----V 1131
Db      805 SRRLAGCPWQGRGYDGLQVLGVIDNQY-----PSVRYHYDGLGSLASRRAGAA 857
Qy      1132 TONFTYDIYGNITACHTTFADGTEHDATFKFANFTDPCQLTEVHTHTPMDNIRLKYDK 1191
Db      858 ASYTRYDAAGN-----RLEGVEFYAR-----EDARQAPAEHELRYSGFSSEVANQ 904
Qy      1192 AG-----RVINITDNHNTENFTYDTLGRV-----QNGQSVYGYDPLNLYSQKT 1237
Db      905 AGECPARWAGNRVRIAGN-----RYRFDALGNLVERIGADGERLRLAYDGAQGLVHLTR 959
Qy      1238 DTLD--CELRYRETMVNEY-----RNG--EMIRLRTGETTIAQGRASKVLLTGDSQ 1287
Db      960 DYADGTRLEAKRYRDALSRRIAKVLVDGVEQVRFQMDGDRQCAEAPAEHLRTTVEHPC 1019
Qy      1288 QSVTL-----TSDKONLSQ--EAYSAYGGRKSTANDASILGYNGERADPVSGVTHLGN 1338
Db      1020 GFWPLALEQACEBDEPPELQLRQAFABEGPL-----PAQCVPALGE 1062
Qy      1339 GYRSYDPTLMKFTPTDLSPPGAGINPYSCLDGPIINRSDPSGHLKWA---WTGI 1392
Db      1063 ARIAF-----FHTDH-----LGTPIQLSDERQQLRWQGVDDWRAV 1098
```

## RESULT 15

```
US-09-328-352-5486
/ Sequence 5486, Application US/09328352
/ Patent No. 6562958
/ GENERAL INFORMATION:
/ APPLICANT: Gary L. Breton et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
/ TITLE OF INVENTION: BAUMANNIT FOR DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: GTC99-03PA
/ CURRENT APPLICATION NUMBER: US/09/328,352
/ CURRENT FILING DATE: 1999-06-04
/ NUMBER OF SEQ ID NOS: 8252
/ SEQ ID NO 5486
/ LENGTH: 3290
/ TYPE: PR1
/ ORGANISM: Acinetobacter baumannii
US-09-328-352-5486
```

```
Query Match      2.5%; Score 222.5; DB 2; Length 3290;
Beet Local Similarity 18.8%; Pred. No. 5.3e-07;
Matches 370; Conservative 249; Mismatches 656; Indels 697; Gaps 101;

Qy      12 ITMSDNNF-----FTQANNFTSAVSG-----GVDPRT 39
Db      459 LTLSLNDNTGVSANDQITQDKNFNKLKSGQETGQVTVLSTDESKTQWETTVAKQDLAD 516
Qy      40 GLVNIQITLGHVNGN-----LGPLPLT--LSYSPINKTIDIGFIGNFGLSVYD 89
Db      519 GIYQKAVVTVAGNISSETAIQKVVDVTTAQAQKLTLSLNDT-----GVSATD 568
Qy      90 ---RKS-----LISLSTGE-----NYKVIETDKTVKQOKKLDLREKDKLXENCVRII 136
Db      569 QITDONSFTLKVQPIVIGEQALLDHYEVSXDESKTQWETTVAD---QKDLADIYO-- 622
Qy      137 HKSGLIEVLTGFNNNAFLKVPKLLNPAGHAIYIDMNFATQPIRLNIYDLDLGHDIPL 196
Db      623 YKA---VITDLAGNISSESAIKQVVDNS-----LWVESTTVIVKPIITD----- 663
Qy      197 LNLVEYGLIKTITLTFPQCKEGYRTFLRPLNRQLNSIHNFSLGNNP---LIMSFGYTP 252
Db      664 ---NTISLVEKQVVISIRLEIANLPTDLS---SLTSVNTTLLENVTYVNFHDE 710
```

```
Qy      253 IGRNGILGOWIT-----SMTAPGGLKETVANYSNNGHHFPOSAN 292
Db      711 VTO-----EWVIEIRPAEFLMSVPEPQTNISIDISLTQAGNTALITHTON---YNVDPTR 762
Qy      293 LPVLPLYVTLMKQVPGAGAPALQAEYSYTHSNYVGGSGNGIWNKLDNLGLMTBYNGST 352
Db      763 SPTLDSLT--FNNIDG-----LISGSAKSGKVIDYNNKGDWMASTIT----- 804
Qy      353 ESRRYDKGKH-----DQIVRIERTYNNYHLLTSCKQONGTI--QTTEAYAYAI-- 400
Db      805 ---NEBKFTLQDLSINTNGEYVAVANTRY-----SENSISGLVTEVPALISIT 851
Qy      401 ---IGHNFDPSQPOFOLPKTKETWRSADNSYSEIETTFDESG----- 442
Db      852 RISPEGVISGATREGSHFY-----KDQGNILIOEPNSNVFDSGIRFPFVMAUGEY 903
Qy      443 ---NEPLTVIKDKTKQIKISP-----STHWEY---YPPA-----GEVD 474
Db      904 RPFILSDQPL-----EEGAQIILISTDKONISGHPQYTADYPAVFLRTPORDISGETL 958
Qy      475 NC--PEPPIGYTRFVK---KIOTPYDSEFKDDEKFIQVR--YSLISQSHVTLKIEGRHY 529
Db      959 SYVHNEPNSFIRAFSGEGNLIATGF-----TDQGFASLQVFPQLKEGFTVQVVDKQ 1013
Qy      530 SATQL-----NSTLFQYNTDKSELGRLLKQTECTKGE 562
Db      1014 NTSETLIEVNFPAVIPHERITQEGLSGVAEONSTIYVADAGNELGKV----- 1063
Qy      563 NGKTVSVHKEFTYTKODDTLQOSHSTTHDNFTIHSQVRSRTG-----R 608
Db      1064 ---TLGDD---NSWSPFSHPSLVNPLIDGKISVQIINDKGLMSPEON 1107
Qy      609 LPEDTDKQIVTQMSYDKGRLLTRTLNSGTPYANTLTVD-----YEANLQDD--- 657
Db      1108 IIVDLPPAPTEINFEDAGDLV---YGAHPESEILVKDQCGNILKMTNNMTDSSGS 1164
Qy      658 ---NRPPFVITTDVANGN-----OLRN-----EFDG-----AGRHV 685
Db      1165 FSTIELGFLNLAETVYVATLADVANGNSLMAQIAPNPAFAPYVDSFTSDGVISQAEINNS 1224
Qy      686 SQCLKSDSGDKFTYTHITQOYDQGRHHTSYSDYLTNGEQCTDPDKVHLSMSKSYNNWG 745
Db      1225 TLVVKDAGD---VVAEIKYGEIDNGWNGSSYFK--LQDLDPVLVDGGEFPLSK---DARG 1276
Qy      746 QIANTHYSYGSEKTIPTDITLT--ATKQLQSNNSNVQTS--KEVYTYPPSQOPIQITLFE 803
Db      1277 QVS-----ADVTYVADTVAPTPASMLVFSDESGSYLVGAELNT-----TIQVFDH 1321
Qy      804 AGHLQSC--HTLTFDGDWRVAKETDAIGQCTIY-----QYDNYNRVLIQ 844
Db      1322 NGQLVNIMNNTIIND-----GTFTIYLSGNNLHGEAFVTVKQDAGNVSEALS 1369
Qy      845 ITLPDGTIVNRKTAPESTDTLIDIRVNGISLGQ-----TFDGLSRLTQSDGGRV 896
Db      1370 INAPLDI---APNPIKNIILD--ANGQNFQAQEAANSQIEVFDSIG---NOTG--- 1415
Qy      897 WATYASAGNQCSPSTVTPDQGFTH--YOQOPELDAVLQVNASNE-----ITQOFSYN 947
Db      1416 WGSTDSAGN-----AGSGFQYTLHGEELTFVVIDRAGNRSIERKQNALIDTTPAPN 1466
Qy      948 PV-----TGALLKAVAGQSLTPIYYPGRLMENINDMKNSYLTMTLGLNGYTDL 1000
Db      1467 PIENIIFENNGQSTFAQAEAGSSIDVDQGT-----NKIGF-----GYDS 1507
Qy      1001 TGTIQKISRDT--HG-----RVTQIYDSSIKITLTANTDLNLRHSGQVTDLAT 1045
Db      1508 SGNVSGFYQGYLLHGEELTFVVIDRAGNRSLEVAQSA---LNDVAPNPPIENIILDD--N 1562
Qy      1046 GHMLT-----TTFEPGLNRE--IGRKLCDSGHTLD--IQOSWLTQOOLANRIVYLGAVL 1097
Db      1563 GQNFQAQEAANSQIEIKNNNGDVVGYSADSAAGNVSGTLYVYVHLHGEELTIVVDRAG-- 1620
```

```

QY 1098 ORTEOYSYDSRNLNOYK-----CDGACPTDKYXHSIVTQNF-----1136
Db 1621 -----NSTEVKQNALIDDIAPNPIENIVLDINGONFTAOAEANTQIEVKNAV 1669
QY 1137 -----YDIYGNITA-CHTTFADGTEDHATFKFAN-----PTDPCQTEVHTHPDMPD 1183
Db 1670 EIVSGYVDAGNVSGYLYQYVYLHGBE--LTFVVVDAGNRSTEVKQNALIDDIAPNPIE 1727
QY 1184 NIKLTKYKAGRVNITNHTGNTENFT-----YDTIGRLONGOGSYGYDPLANRLVSQKTD 1239
Db 1728 NIL------DANGONFTAOAEANTQIEVKNAAGEVIG-----SGSTDS 1765
QY 1240 L-DCELYRETMLVNEVRNGEMIRL--RTEGTIAQORASKVLTGTSQOSVILTSD 1295
Db 1766 MGNVSGTFYQVYL-----HGBELTFVVVDAGNRSTEVKQNALIDDIAPNPIENIIPNEN 1820
QY 1296 KONLSQEA-YSAYGKHKSTANDASILGYNGERADPVSG-----VTHLGN 1338
Db 1821 GONFTAOAEANSKVEKNAAGEVVGSGY-VDSAGNVSGYLNQYVYKGBELTFVVIDQAGN 1879
QY 1339 GYRSYDPTLKRFPDLSPPGAGI-----NPYSYCLGDP--INRSDPSGHL-----1384
Db 1880 --RSIEVKQTAFLDNTAPENATNLVSEBGSYLSGMAEPNATIQIFDQYQQLNOMN 1935
QY 1385 --SMAWTGIGWG--TAGLLLTATGMAIAAGGIAAIASTTALAAGALSVTSDI 1439
Db 1936 NVNWDGTFNITLNSNTYHGEVFKV---VVDDQGNLSGEVT-----1973
QY 1440 TSIVSGALBDASPVSISILGVMGMAAGIABSAIKGTKLATHLGAFAEDGENALLKS 1499
Db 1974 ---VKAPLDDIAPVPAASDLVFNEDGSSLSGVAEP-----NTFIQIFDONGQO--MNT 2020
QY 1500 TSESSRIKMGVTRSLSDEIYVRNEGOYIKHRSRGYTDNFMKGEOALVHGDKDGFLYHT 1559
Db 2021 WSQSVNADGTFIIFGTYNLHGEFTVI-----VKDLAGNVSEAVSV-----2062
QY 1560 EGNKHNKGFPYTRHTPEQLVDYLDKNNIIVDLTQGD--KPVHLLSCYKSSG 1609
Db 2063 -----KAPLDDIAPNPIKNIYVDANGOSFTAOAEANSQIEIFDSFGSQIG 2107

```

Search completed: January 30, 2006, 09:50:44  
 Job time : 46.5112 secs

**THIS PAGE LEFT BLANK**

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: January 30, 2006, 09:43:19 ; Search time 104.198 Seconds  
(without alignments)  
6708.668 Million cell updates/sec

Title: US-09-889-874A-23

Sequence: 1 VIKKFLKLRRTITMSNNNF.....PKIIIGRTKTKYKTPRP 1673

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA Main:\*

- 1: /cgn2\_6/prodata/1/pubppa/us07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/prodata/1/pubppa/us08\_PUBCOMB.pep:\*
- 3: /cgn2\_6/prodata/1/pubppa/us09\_PUBCOMB.pep:\*
- 4: /cgn2\_6/prodata/1/pubppa/us10A\_PUBCOMB.pep:\*
- 5: /cgn2\_6/prodata/1/pubppa/us10B\_PUBCOMB.pep:\*
- 6: /cgn2\_6/prodata/1/pubppa/us11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	932.5	10.5	982	US-10-365-742-106	Sequence 106, App
2	517.5	5.8	1317	US-10-369-493-11243	Sequence 11243, A
3	420.5	4.7	2364	US-10-156-761-7834	Sequence 7834, Ap
4	395	4.4	820	US-10-156-761-7990	Sequence 7990, Ap
5	392	4.4	1250	US-10-156-761-7572	Sequence 7572, Ap
6	391.5	4.4	2386	US-10-156-761-7751	Sequence 7751, Ap
7	386	4.3	2234	US-10-282-122A-46565	Sequence 46565, A
8	361	4.1	1510	US-10-418-861B-55	Sequence 55, Appl
9	353.5	4.0	843	US-10-282-122A-76214	Sequence 76214, A
10	350	3.9	1515	US-10-282-122A-47600	Sequence 47600, A
11	342	3.9	1397	US-10-282-122A-43059	Sequence 43059, A
12	340.5	3.8	1411	US-10-282-122A-43060	Sequence 43060, A
13	339	3.8	1426	US-09-912-020-340	Sequence 340, App
14	339	3.8	1426	US-10-282-122A-42617	Sequence 42617, A
15	339	3.8	1426	US-10-282-122A-44588	Sequence 44588, A
16	331.5	3.7	1377	US-09-815-242-10384	Sequence 10384, A
17	331.5	3.7	1377	US-10-287-274-467	Sequence 467, App
18	331.5	3.7	1377	US-10-282-122A-42731	Sequence 42731, A
19	328	3.7	1582	US-10-282-122A-69415	Sequence 69415, A
20	322	3.6	1385	US-10-282-122A-68242	Sequence 68242, A
21	317.5	3.6	1565	US-10-282-122A-44588	Sequence 44588, A
22	314	3.5	1395	US-10-282-122A-50586	Sequence 50586, A
23	307.5	3.5	932	US-10-282-122A-44617	Sequence 44617, A
24	306.5	3.5	1530	US-10-282-122A-68175	Sequence 68175, A
25	298.5	3.4	1531	US-10-282-122A-49308	Sequence 49308, A
26	297	3.3	1512	US-10-282-122A-78386	Sequence 78386, A
27	291.5	3.3	1253	US-10-282-122A-68914	Sequence 68914, A

28	291	3.3	2346	US-10-072-012-491	Sequence 491, App
29	290.5	3.3	2802	US-09-808-602-81	Sequence 81, Appl
30	290.5	3.3	2802	US-09-800-198-69	Sequence 69, Appl
31	290.5	3.3	2802	US-10-072-012-489	Sequence 489, Appl
32	289.5	3.3	1364	US-10-282-122A-75309	Sequence 75309, A
33	285	3.2	2613	US-10-038-854-42	Sequence 42, Appl
34	285	3.2	2628	US-10-038-854-40	Sequence 40, Appl
35	285	3.2	2721	US-10-038-854-38	Sequence 38, Appl
36	285	3.2	2725	US-10-038-854-36	Sequence 36, Appl
37	284.5	3.2	2715	US-10-042-865-52	Sequence 52, Appl
38	284.5	3.2	2715	US-10-029-020-51	Sequence 51, Appl
39	282.5	3.2	1769	US-10-723-860-4494	Sequence 4494, Ap
40	280	3.2	1317	US-09-815-242-5118	Sequence 5118, Ap
41	280	3.2	1317	US-10-282-122A-43495	Sequence 43495, A
42	279.5	3.1	944	US-10-282-122A-47806	Sequence 47806, A
43	278	3.1	985	US-10-282-122A-43061	Sequence 43061, A
44	276.5	3.1	1438	US-10-282-122A-78360	Sequence 78360, A
45	274.5	3.1	2144	US-10-723-860-2303	Sequence 2303, Ap

## ALIGNMENTS

RESULT 1  
US-10-365-742-106  
; Sequence 106, Application US/10365742  
; Publication No. US20030204868A1  
; GENERAL INFORMATION:  
; APPLICANT: Collmer, Alan  
; APPLICANT: Alfano, James R.  
; APPLICANT: Carlinhour, Samuel W.  
; APPLICANT: Schneider, David J.  
; APPLICANT: Tang, Xiaoyan  
; TITLE OF INVENTION: PSEUDOMONAS AVR AND HOP PROTEINS, THEIR ENCODING  
; FILE REFERENCE: 19603/4112  
; CURRENT APPLICATION NUMBER: US/10/365,742  
; CURRENT FILING DATE: 2002-02-12  
; PRIOR APPLICATION NUMBER: 60/356,408  
; PRIOR FILING DATE: 2002-02-12  
; PRIOR APPLICATION NUMBER: 60/380,185  
; PRIOR FILING DATE: 2002-05-10  
; NUMBER OF SEQ ID NOS: 209  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 106  
; LENGTH: 982  
; TYPE: PRT  
; ORGANISM: Pseudomonas syringae pv. tomato DC3000  
US-10-365-742-106  
Query Match 10.5%; Score 932.5; DB 4; Length 982;  
Best Local Similarity 29.7%; Pred. No. 7,8e-56;  
Matches 285; Conservative 142; Mismatches 357; Indels 175; Gaps 33;  
642 ANTLYDYELNMLQDNRPPVTTTVDVNGOLRNFEGDAGRHSQCLNDSDDGKFFYT 701  
31 SSTLGRYD-----DNR--CTTTDDN-VQYVESDPIGSHVK-----GPIQKT 74  
702 HTQOYDEQGRHHTSTVDYLTNGRQOTDPDKVHLMSKSYDNWGQJANTHMSYGSBKIT 761  
75 WKQSGPBER-----ISGRSET-----WLN-----FGKEDRIR 103  
QY 762 VDPITLTATKQLOLSNNVQTKGKVTTPSOQPIITLFDEAGHLOSCHTLTRDGDV 821  
DB 104 ---TLTACKTGRSHRTHSMRSANLTT---TEOEL-----SRQFLYDGLGRC 144  
QY 822 RKETDAIGCTIYYQNYNRVIOITLPDGTIVRKAPSTDTLLDIRV---NGIS--- 875  
DB 145 TEGRDLQOSTLFSTYDNMRMSSTILADSVINRSTAPQSSSLATMLBHVHNGTTRIV 204  
QY 876 LQOQTFDGLSRLLQSGDGRWVAAYTYSAGNDQCPSTVITPDQGFHYQYQPELDVAI 935  
DB 205 AGQKFDGLERVTQTKGDRVREQFNVDAGBMQ--PRRTTAGLDNINFTYTRALTDQIFSS 263

QY 936 ASNETTQPSYNPVTGALLKAVAGSGSLPIYPSGSLNMENTINDMKMSYTW-----T 989  
DB 264 TADETKAFDYDKTSARLIBATNPQGRTRYVDVHNLTEBTMDNL--LQAMETRRQSS 321  
QY 990 LRGLNGEYTL-----TGTIOKISRDPHGRVTOIKDSIKTTLNAYDLNHHISQVTDLA 1044  
DB 322 LIGRPYKRTDLKGAAGAETRYDYDYLGRIRPINSNLRTTIDYDVLGQLCKYATDLO 381  
QY 1045 TGHMLTTVEPDGIRREIGRKLCSGSHLLDIQGSWLKTOQLANRIYKLVNGVLOFTEOYS 1104  
DB 382 AGGVVILIDMEYDQGGELIRLTQTASNOAALLTQTWAVDGLIKRDLQOAGSPILHETFS 441  
QY 1105 YDSNRLNOKKCGAEPTDKYGSITYQNPVTYIYGNITACHTFADGSDHATFEKAN 1164  
DB 442 YDPRGRLLTVNYLSSLPRELQREMRQIFSPDELNITLCQRFRTDGTGSERAAFYKGS 501  
QY 1165 P-----TDPCOLTEVHHTHPMPDNIRLKYDKAGRVINITDNHNTENFTYDTLGRL--- 1216  
DB 502 PGDDKHDKRQQLSIATTPRKTPDPFTSYDANGN--QLKDEHGN--SLHYDSQSRLLQY 557  
QY 1217 -QNGQG--SVYGYDPLNRLVSQKTDIJDCEI--YRETMLVNEVRNGEMIRLRTGETTI 1271  
DB 558 AETGAPISQYRYDGHQVLVATR--DGNSEILRFYEGHQLSSTVQEDQRTQYHLGEOPL 616  
QY 1272 AQC--RASKYLLTGTDSQGSVILTSQKONLSQEAAYAGKHSTANDASILGNGERAD 1328  
DB 617 GQQVLDVAEQTLLLTLDNOSVWGEFOGQRLKAVSYSGERHSEBALLSAGNGEYRE 676  
QY 1329 PVSGVTHLNGYRSYDPTLRFHTPDSLPFGAGINPYSCLDPIINRSDPSGH----- 1383  
DB 677 AANRWYLLNGYRKYRNPILMRFHSPDLPFAEGVNPYTYCLENPLIALNRPTHDSGQ 736  
QY 1384 -----LSNQ-----ANTGIMGTA-----GILLTIANGMAIAAGGI----- 1416  
DB 737 TGRLLRPDEGALPMQGGSGDIMGWVGIGVFTVLGVAATIAATIGATPVGTVLGI 796  
QY 1417 -----AAIASTSTTALAFG-----ALSVTSIDITSYGALEDAKSPKASILGNVS----- 1462  
DB 797 SMTASAAAASVSTIGALIVTALTASTANTVAIVNN-----DQNGEYGVGMLGIAAV 851  
QY 1463 -----MGMAAGIAESAIKGTGTLA-----THGAFAEDGENALLKSTSESSR 1505  
DB 852 PVGLVGRGAGVAVARAFAAKVAANAAGTIGVRSVRIG--LAAAGARTISSAASSAR 909

RESULT 2  
US-10-369-493-11243  
; Sequence 11243, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OR INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 11243  
; LENGTH: 1317  
; TYPE: PRT  
; ORGANISM: Methanobarcina maei  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(1317)  
; OTHER INFORMATION: unsure at all Xaa locations  
; US-10-369-493-11243

Query Match 5.8%; Score 517.5; DB 4; Length 1317;  
Best Local Similarity 21.2%; Pred. No. 1.5e-26;  
Matches 334; Conservative 194; Mismatches 505; Indels 541; Gaps 72;  
QY 39 TGLYNIQITLGHVNGNLGPTLPLTL--SYSPANKTDIGFGIG--FNGELSYVDRKNSL 94  
DB 56 TSGYFYQYDLSIPRGG-----LPLTVSRSYNSMDNRSGLFSGGWTFFNYMNLTVDNNGN 110  
QY 95 LSLSTGENYK--VIETDKTVKLQCKLDNLRPEKOLKENCYIILKSGSIEVLTGKNN 151  
DB 111 VYVAGDGHDTYILMPDGYRSRPLSVFDDL-----KNSDetyTLTKDQT 157  
QY 152 APLKVPKLLN---DAGHAYIDMNFPEATOPRLNRIYDDLGDHIDPLNLRYOGLIKTI 208  
DB 158 KINPSEBGLVNLVDKNGNQI---NFTYGBQUTKY--TDSAREL--ILAYDNG----- 206  
QY 209 LTLFPGQKEGYRTELFLNRQLNSIHNFSLGNENPLTWSPGYTPIG---KNGILGQ-- 261  
DB 207 -----HISITDPMGRVWSYSDQGNLIQCKNPIGKLS 241  
QY 262 -----WTSMTAPBG--LKETVYNSNNQGHHPQSANLVLVYVTLMKQVPGAGOP 311  
DB 242 YTYDENHMTSITDPPGNHMKNTY---DEKGVISQS----- 276  
QY 312 AIQAEYSYTSNHYVGGSGNIGNNKLDNLXGLMTEYNGSTESRRYKDEKGDQVRIER 371  
DB 277 -----NSLNATY---TRNYDS--ENRK----- 293  
QY 372 TYNNHLLTSECKQONGYIQTETAYAYAILGNHPSQPSQFOLPKYKTETRWSADNSYRS 431  
DB 294 -----TTETD-----PFGKNKYTSEHEFW-- 313  
QY 432 EITET-----TPDESGNPLKVIKDKKQKIISPTHWMEYYPAGVDNCPREPYG 482  
DB 314 ELNETYQGLYTSIYADENGNRISVTNENSKTKLA-----YDANGYIIR--TINPLG 364  
QY 483 FTRFVKKIITPYDSEPKDPERKIQYRYSILSGSSHVTLKIEBRHYSATQLNS----- 537  
DB 365 YSK-----SMTYDS-----KNNLIQTDLBRKTSREYDNNLSISDALG 406  
QY 538 --TLFOYNTDKSELKLLKQTECK-----GENK---TYSVHAKFT 574  
DB 407 HEIVFSYD---KXGVYIGETDSNKKTATFSYNNNGDQITITDANGKTSAFYDTYGVKVT 462  
QY 575 -----YTKQDDTLQOSHSTTHDNFTIHSQVRSRTGR-----FSDTDRKDI 618  
DB 463 TKTDAKGNRYTFQYDALDNLISITD-----PMGQTTSTNMLEIYKLVLMKLVKQSI 516  
QY 619 VTQMSYDKLGRLLRRLNSGT---PYA-----NTLYVDEYELNNLODDNRPPVIT 665  
DB 517 LITLXISLKKRMQWVXGLGTNNMPLALWPRQTKHKTISYDDPLNRQ-----VS 568  
QY 666 TTDVNGNQLNEPFDAGRHSV-----QCLKSDGCKEYTTHTPOYDE 708  
DB 569 VTNALGKTRNKYDAIGNKISITNAYGKSTRYSYNSLQVLYKVNAMGK---VVRVYUDA 625  
QY 709 QGRHHTSYD-----YLTNGRQCTDPDKVHLMSKSYDNMQILNTHMSYGVSEKIT 761  
DB 626 VGNLISITDENGKINYGYSLNQVSVTDALRKTKTRNKDVAENKISITNAYGKSTRYS 685  
QY 762 VDPITTLATKQLQSSNNVQTKGKVTYTPSQQPIQITLPEAGHLQSCHTLTREDGMDRV 821  
DB 686 Y-----NSLN-----QLVAVTDAMGVAVRYNDVAGNLIS----- 715  
QY 822 RKETDAIGQCTIYOYDYNRYIQTLPDGTIVNRYAPFSTDPLITDIRVNGISLGQOTF 881  
DB 716 --TTDANGRKKNYGYDLSNRQVSTINALGKTRNKRYAVANGKISSITANMR---LTKYSY 770  
QY 882 DGLSRLTQSDG--GRWAYATSA--GNDQCPSTVITPD--GQFINQY-----OPELDAY 932  
DB 771 DSNRLVKVTDAMGVAVRYTDAVGN-----LISTTAKGHKTDYGDSDLRQVSIIDPL 825





Db 1608 T-----TTT-----LNKVSERQKDT-----YTYDLAAGLTELREASGQ 1642  
Qy 1097 LQREQSYSTRNL-NQYK-----CDABCEPTDKGHSIVQNTFYDIIGNITACHT 1148  
Db 1643 TAOSQCELYDQALTAHYHTTTTGICADKTKTASDEKGTAPYQATYTYRLGLOSITN 1702  
Qy 1149 TPAAGT-----EDHATFKFANPTDECOLTEV-HHTHPMPNIRLKYDKARVI 1196  
Db 1703 TDSAGTATLDYLYPGVDDTGTTANANOPHGRKIDHKTGSTTSDDLYTDDGTMK 1762  
Qy 1197 NITDNHNTENFYDVLGRLO-----NGQGSV---YGYDP---LNLVLSQKT---DTL 1240  
Db 1763 QRYV-PGTTIDYTYTPQGLEAVKTKSSGSELTRYVADAGNILVTTTQETVASIDGM 1821  
Qy 1241 DCELYEETMLVNEV-RNGEMIRLLRTGETTIAQORASKVLLTGTDSQOSVILTSDKQN- 1298  
Db 1822 ELRTNGTYYATRVASGATVAMRTTEGTTAN--GKVTYLMADYQASTQALVADASTG 1879  
Qy 1299 -LSQEAYSAYGKHKS-TANDASILGYNGERADPVSGVTHLNGYRSYDPTLMRFTPDSTL 1356  
Db 1880 ASTRRRTTPEDEBSGTLPTGTONGFLKGTEDTSTGLSL--GARAYDPNLRGRLSPDPL 1937  
Qy 1357 -SPEGAGINPYSCLDGPIINRSDPSGHLISWQAWTIGMGIAQ 1398  
Db 1938 ATPAPQMLNAYSATNNPISYDPSG-LCRDIDCGGYPAVG 1979  
  
RESULT 4  
US-10-156-761-7990  
; Sequence 7990, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156, 761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 7990  
; LENGTH: 820  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-7990  
  
Query Match 4.4%; Score 395; DB 4; Length 820;  
Best Local Similarity 23.8%; Pred. No. 2,7e-18;  
Matches 242; Conservative 120; Mismatches 375; Indels 280; Gaps 50;  
  
Qy 501 DDEKFIQRYSLISQSHVTKIEBRYSATQLINSLFYQYNDKSLGRLKOTEC-T 559  
Db 26 DTEGTFYFRYDAMGRVVKIT-----TAEGRVVFTYD-DANVTSMRLRTGFNS 74  
Qy 560 KGENGKTVSVHAKFYTKQDDTLQOHSITTHDNFTIHSQVRSYRQLRSDPTKQIV 619  
Db 75 DGHTEPTV-----TYAVTSBSVTAAGTTATDPTT-HATKQHDSDQVSDVTDAMGK 127  
Qy 620 TOMSYDKLGRLLTRT-LNSGTPYANTLLTYDEL-NNLQDNNRPFFVITTTDVGNGQLRN 676  
Db 128 RSTKDNASHIDISTIDAMSGSTTPENAVTDYGFNTNNLE-----TTQPGGKTVN 178  
Qy 677 EPDG-AGNHVSQCLKDSGDKGFYTIHQYDQGRHHTSYSDYLTNGRQGTDPDKVHL 735  
Db 179 HMQGTAGADVPKDSINPDGEKTDFT-----YDAVGN-----TM 211

Qy 736 SMKSVDNNQOLANTHMSYGVSEKITVDPITLTA---TKOLQSNNNVQGEKVTYTPS 792  
Db 212 SVAQTGTGGNNVYTT-----NPAFPCGGFAGRCREKTKMTAKVTT----- 256  
Qy 793 QQPIQITLPEAGHLOSCH-----TLTRDGMVRKETAIDIGQCTIYOYNNY-RV 842  
Db 257 -----TFHDSAGNLDTVPAPRLAKITYYDALGKTKTVDARGVTVVTYNNRDRTRI 311  
Qy 843 IQITLP-----DGTIVNRKAPFSDTLTDIRVANGISLQGTDFDLSRLT-QSQ 891  
Db 312 VDTTNRAVEYWDGNGNLQR-----TDSGTITRYD-----FDPLORETLRTLO 356  
Qy 892 DGR-VWATYTSAGNDQCEPSTVITPDGQFIHYOYBELDAVLQVANSNETQOFSYNPVT 950  
Db 357 DSGQTLIAY-----TPSGNDVYQ-----DPAQ-----TVDYTNNEV- 388  
Qy 951 GALLKAVAEQSILPIYPSGRK--MENINDMKRNSYLTWTLNGLNGYTDLTGTIOKIS 1008  
Db 389 -----NKLAEKDPAGRVTTYKYNNDVRYTT---TYPG-----GTQOKVD 426  
Qy 1009 RDTHGRVTOIKDSIKTT-----LNYDDLNRHIGSOV---TDLATGHMLTTTVEFDGL 1058  
Db 427 PDNSSRPTKITKTSFGKTPVDLAYSYGYGTSGTTEGSKTSSIDNVTG-LKITYYDGA 484  
Qy 1059 NREIGRLCDSSGHTLIDIOQSWLKTQOLANRIYKLVGLQRTQOYSYDSHRNLNQYKCDG 1118  
Db 485 GRFSYAE-EKKGITLNL--SSWLYCYDLAENL-----TSGTAAAGCGATTYTVND 531  
Qy 1119 ACEPTDKYHSIVTONFTYDIYGNITACHTPADGTEHDATFKFANPTDECOLTEVHHTH 1178  
Db 532 AQOLTAENGSS---TTWMSYDKIGNETAGAST-DEGRTAEKW-----TDSOLTSL- 578  
Qy 1179 PDMPDNRLKYDKAGVININTDNHNTENFTYTLG--RLONGGSVYGYDPLNRLVSQK 1236  
Db 579 -----TWGK--TYTGQYSTDQSERIRLGGTYTHANGPIGAG----- 614  
Qy 1237 TDTLCELYRETMVNEVANGEMIRLLRTGETTIAQORASKVLLTGTDSQOSVILTSDK 1296  
Db 615 TSTAGVDYGFNR-----EPGGTILMSNTGGKAYCYVLTALGSLVLTDE 658  
Qy 1297 -QNLQEAYSAYGKHKSANDASILGYN--GERADPVSGVTHLNGYRSYDPTLMRFT 1352  
Db 659 TGAQVNTYVAYSPRGVTAASSTSEKRPQYRPAAGYOD-VTGLYHGARY--YDPNIGFNS 715  
Qy 1353 PDLSPFGAGINPYSCLDGPIINRSDPSGHLISWQAWTIGMGIAGLLTATGMAIAA 1412  
Db 716 PD--PESQER-NBYLYABGDPVNRIDPNGLFESDALDTGSDIFGVVYTGCLAGVTLAAE 771  
Qy 1413 AGGIAAIASTSTTALAFCALSVTSIDTSIVSGALIEDASPRASSILGWSMGMAAG 1469  
Db 772 TGSIAVAAA-----VGVVAGVGSVAGVGLAVGSCALG-GAAG 810  
  
RESULT 5  
US-10-156-761-7572  
; Sequence 7572, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156, 761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697

PRIOR FILING DATE: 2001-08-02  
 NUMBER OF SEQ ID NOS: 15109  
 SEQ ID NO 7572  
 LENGTH: 1250  
 TYPE: PRT  
 ORGANISM: Streptomyces avermitilis  
 US-10-156-761-7572

Query Match 4.4%; Score 392; DB 4; Length 1250;  
 Best Local Similarity 21.2%; Pred. No. 8.7e-16;  
 Matches 271; Conservative 187; Mismatches 452; Indels 368; Gaps 66;

254 GKNGILGOWITSMAPGKLTETVYNNNOGHHPQ-----SANTPLVPLVTLMKQVPG 307  
 DB GMEVBAOGLRK-TANAABEDASROKAKICEKDPVDVATGRNVLPQI-----DVSL 113  
 308 AGOPALAEYSYSHANYGSGNSGNIWNNKLDNLY-----GLMTEYNGSTES 354  
 DB PGQLPLVVKQPFESYRLGGWFPWSTLDQRLBVDVAGVVLVGEDQLVLAIPH-----PA 170  
 355 RRYKDKGHQIVRIERTYNNHLLTSECQKQNGYIQTETAYYAIIGANFDSQSPQL 414  
 DB PGVPLRSHGRWMLDRIDGGYTL-----TDQKSGHIR----- 203  
 415 PKTKETWRSADNSYRSEITETFEBSGNPLTKVYIKDKTKQKIIISPTMWEYYPAGEVD 474  
 DB -----HVVNSDPLAVALBQLDRNGNMIT-----FEY-----DAD 233  
 475 NCP---PEPYGTFTRVK-----KIOTPYDSEFKDPERK-FIOYRSLISGSHVTLKIEE 526  
 DB GAPRYLTHSGGYRLRISTEAGRVTAHLAASAVDGGQDLIRYV-----TDGHL----- 283  
 527 RHYSATQLANST--LFOYNTDKSELRLKOTE-----CTKGENGKTVSV 569  
 DB -----TEVNSSGRLQYVCD--ELGRITSWTDNDSHFSYAYDEDERCTH--GSGAAGHL 335  
 570 VHKETVYKOD-----DTLOQSHSITTHDNFTIHSQVSRYGRLESPDPTDIYTM 622  
 DB RSTPFAVANDPBGATATTTVDSIGQTH--YLNR-----RCQYIAEDALGAVTRY 386  
 623 SYDKGHLTRTLNSGTPYANTLYDY--ELNNLODDNRPPFVITTTTVNGNQLNEPDGA 681  
 DB ORDRYNLLSQT--DPLGHTTSFRYDDAGNLVAATRP-----DGBABAEYNAL 433  
 682 GRHVSQCLAKSDGCKYTHITQOYDECGRHHSTYSDYLTNGRQOQDPDKVHLSMSKSY 741  
 DB GLPKK--LVNPDG-----TTRQTFDERG-----NLTSVDPF--GQTFRGY 472  
 742 DNMGOIANTHMSYGVSEKITVD-----PITLT--ATKQLOSNNNVQTKGVTTYPSPQ 793  
 DB DEGGRLTSMTDPLGHTTGIVCDRAGLPLVTDPDLAGAVTRYERDAFGRTTA--ITDPTGAT 530  
 794 QPQIOTLTFDEAGHLQ-----SCHTLTRDGMHVRKETAIGQCTIYQDYNRVIT 846  
 DB TRLEMTV---EGHLSRRTAPRGSTESMTYDSENGCTHTDPVGVSLFEYTHPLPRTART 587  
 847 LPDGTIVNRKAPRSTLTLDIVNGISLGGQTFDGLSLRTSGODGRWVAITYSGND 906  
 DB GPDGV---RY-BFEND--TELR-----LSQVTPNH--GLTWNVAYDAAGR 624  
 907 QCSSTVITPQGFHYOYOPELDAVLQVAS--NEITQPSY-NPVTGALKKAVAEQSL 963  
 DB LVAEITP--DNRITLYEY--DPAGLMSRTNALGGMIAFENNELGQYIKDAQOAT 677  
 964 TPVYPSGRKLMENINMKMSYLMTLRGLNGYTDLTGTLQKISRTHGKVTQIKOSI 1023  
 DB TVAYDFTDQLAQA-----TSPDGTALTLTLRBHGMSEAVNGR 716  
 1024 KTLIANDDLNRHT-----GGOVT---DLA-----TGHMLTTTVEVDGLANREIQRKLC 1068  
 DB ALTYDYDELGKRTTRTPSSATITMSYDVAGRTSMGSGSIDFVVDDEGREIGRRP-- 774  
 1069 SSGHTLIDQSWLKTQQLAARIVLNGVLTQRTBOYSYDSNRRLNQYCDGABECTDKYGH 1128

DB --GH-ITLHSPDALGRITSGSV-LGPAGRTDHRQ-----TYRAD-----GH 814  
 775 SIYTON-----FTYDIYNTACTH-----TF--ADGTEHATFKFANP----- 1165  
 DB LIHIEDQLSGRRFDDLAGRTAVAHAVNMTETAYADAVNGQSASAPAGPGEAIGNR 874  
 1166 -----TDFCQLEVHNTHFPM-----PDNIRLKYDKAGVINITDNHNTEN 1207  
 DB TYTGSLITRAGVRYEBIDGRIILQKRLSKRPDTRWTWAEDELTSVTPPDGTRMR 934  
 1208 FTYDTLGR-----LONGGSVYGYDPLNRLVYSOKTDTDCELYRETLVNEVR--N 1257  
 DB YTTDPLGRKTRAKRLAEDSGTIV-----ERVDFMTDGTVLCEQTTSPLPHQVTLTWD 988  
 1258 GEMIRLLRTGETTIIAQGRASKVL-----LVTGDSQQSIVLTSDKNLSQEAAYSAY 1307  
 DB HOGRLPVTQTERIVAADAPOQEIDSRFALVTDLVGPSB-----LDERGSIWTRATLW 1045  
 1308 GKXSTFANDAST--LGNGSRADPVSGVTHLGNGYSYDPTLRFHTPPDSLPFGAG-GI 1364  
 DB GSTTWAKNSTAYVPLRPPQYDPEGLHY--NFRHYDEPTARVYTPD--PLGLGPAP 1100  
 1365 NPYSYCLGDPINRSDPSG 1382  
 DB 1101 NPAAY-VPNPHMADPLG 1117

# RESULT 6 US-10-156-761-7751

Sequence 7751, Application US/10156761  
 Publication No. US20030119018A1  
 GENERAL INFORMATION:  
 APPLICANT: OMURA, SATOSHI  
 APPLICANT: IKEDA, HARUO  
 APPLICANT: ISHIKAWA, JUN  
 APPLICANT: HORIKAWA, HIROSHI  
 APPLICANT: SHIBA, TADAYOSHI  
 APPLICANT: SAKAKI, YOSHITUKI  
 APPLICANT: HATORI, MASAHIRA  
 TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 FILE REFERENCE: 249-262  
 CURRENT APPLICATION NUMBER: US/10/156,761  
 PRIOR APPLICATION NUMBER: JP 2001-204089  
 PRIOR FILING DATE: 2001-05-30  
 PRIOR APPLICATION NUMBER: JP 2001-272697  
 PRIOR FILING DATE: 2001-08-02  
 NUMBER OF SEQ ID NOS: 15109  
 SEQ ID NO 7751  
 LENGTH: 2386  
 TYPE: PRT  
 ORGANISM: Streptomyces avermitilis  
 US-10-156-761-7751

Query Match 4.4%; Score 391.5; DB 4; Length 2386;  
 Best Local Similarity 20.1%; Pred. No. 2.6e-17;  
 Matches 379; Conservative 214; Mismatches 688; Indels 601; Gaps 93;

163 NP---AGHAIYIWNFEATQPRINRIYDDLDGHDIPLNTLEYQLITLTFPGQKRGY 219  
 DB NPDGCTIKTMLDIYQKT-----YGDGDDIVLPVAINNRKIDL-----DNKVG- 740  
 220 RTBLRFLNRQNSIHNSLGNENPLTWSFG-----YTPIGKNG 257  
 DB SAEIINF--PRIQIHH--DLGATTKVSYGAFANACIDHLPQASNTDQCYQKMTPEBETE 797  
 258 ILGOWITS-NTAAGGLKETIYVNSNNNGHHFPPSALNPLVPYTLTKMQVVG-----AGQ 310  
 DB SKTGWPKKFLVTVKVEDPVP--TTNODG-----APVMTTSYTYEGAGWRFETGD 844  
 311 PAIOAB-----YSYTSNHYVGGSGNGIWNKLDNLYGL-----MTSEYNGSTESRRYKD 359

Db 845 PLVDEBSMDWQYQVYTTGADTGQKTKMYLRYGLDGRSTKIDTSATKTVTVND 904  
 Qy 360 KEGHDQVRIERTNNY-----HLLTSECQONGVIOETTERAAYAILGN---PDSQ 408  
 Db 905 GGG-----NNYADHAMLAGHTLSTLSRDTDI--SHEHTYHTYSHNAQYDGL 951  
 Qy 409 PSCQOLPKTKETWRSADNSYRSEITETTPDES-----GNPLTKYIKCKTKOKIISPS 461  
 Db 952 PDHAFVRESKSTNTTKISGGRREHVENENEYDSEASTTEGLPR---TDDMGSNVS-- 1006  
 Qy 462 THMEYYPAGEVDN-CPEPEYGFTRPVAKKIQTPIYDEHFDKPEKFIQYKSLIGSOSH 520  
 Db 1007 -----DNRC-----TYGRAVNTD-----NYDSTGAORWT 1031  
 Qy 521 TLKIEBHYSATOLNLTLPQYNTD--KSELGRLLKQECTKGNGKTVSVHHTYTKQ 578  
 Db 1032 VVOQVQVHYVVG---CSSIADSNODGYTSTLYDNATSIDANKPVQGNATEY--RTTK- 1084  
 Qy 579 DDTLQOSHISITTHDNFTIHSQVRSRYTGRLLFSDTDTKDVIYQMSYDKLGRLLTRTLNSG 638  
 Db 1085 -----AGRYSSWTG-----YDAGRIVMS--EDG 1107  
 Qy 639 TPYANTLTVD---YELNLODDN-----RPPVY-----TTTDVNGQ 673  
 Db 1108 KHNRSITVSPANTWPMNGITSTSPDPDGTATARGPLTFETWTSFPAATPTTSKDANGNI 1167  
 Qy 674 LRNFPGAGRHVSQCLKDSDDG---KP-YTHHTQO-----YDEQGRHHTSTYS 718  
 Db 1168 TKVTLDAAGREVEYWKPTETGSSPSMKFSYTIPTSTMSAGVPDADGYPHVAHSHLQSGS 1227  
 Qy 719 DYLTN-----GR-----OQTPDKVHLSMSKS---YDNMGOLA----- 748  
 Db 1228 TYLVSHAVTDGLGARARETQTEPLPSVDPATQIVPFRQVAVTRDSAGQVAGASAVFRNQ 1287  
 Qy 749 NTHMSYGV-----EKITVDPITLTKLOLSNNNVQCKERTY-----TP 791  
 Db 1288 GTASGGSPSPQPSDLPSPSYDLVDMAGRIVSSQIQVNGTPQKAGRVDTSLGDTYSVTP 1347  
 Qy 792 SQQPIQITLFPDAG-----HLOSCHTLTRDGW---DRVKEKPDALQCTIYOYDNVR 841  
 Db 1348 VDSRTDTTYTVYQVQSVKVEHTASSAYTTAGYAKELAQITDPKGNNTLYYDMAQ 1407  
 Qy 842 VIOITLPPDGIYNRY---APFSTDTLITDIRVNGISLGQTFPGLSLRLOSQDGG---R 895  
 Db 1408 RKTDDPDAAGLSSSEYNNNGQVSOQTATNTDVOYVLTG---YDNLSTATSVRSGADELA 1464  
 Qy 896 VWAY-----TYSAGNDCPSVIVTPDQFHHYQPELDDAVLQVANSNETQO 943  
 Db 1465 AMVWDDPRAATGKQGITSAVSRDASGNTYTKTKF-----DERGRPLMTTVLP 1514  
 Qy 944 PSYNPVTGALLKAVA--EGOSLTPYYP-SGRLLKMEI---NPMKKSYMTLRLGLENG 996  
 Db 1515 TVVNGIADYTTSTVYDAADHITSYSPAGKLAEXKTYTYDDYGOPTRLTSSIG--- 1570  
 Qy 997 YTDLTGTLQKISRDTGRVTOIKDSIKTTLN-----YDINRHISGO-VTDLATG 1046  
 Db 1571 ---GTAVIDNTYVAGRLVE-RDYGAFFGNGIGIQARQGYDDBSN---GTRMLRSIAT- 1622  
 Qy 1047 HMLTTFVFDGIANRIGRKLCDSSCHTLDIQOSMLKTOQLANRYKLVKNGVLOREHYQYD 1106  
 Db 1623 ---TTINDLVSEAOKOTYLLVDNTGKLELRB-----QASGQTAQSCLEMYD 1666  
 Qy 1107 SRNRLN-QYK-----CDGACPTDKYGHISIVTQNFYDIYGNITAGHTTFADGT----- 1154  
 Db 1667 DOSRLTLAYHTTAAACADTTKTSIDFKGTSPOYGYVYDRLGNQSVTDSAGAAATR 1726  
 Qy 1155 -----EDHATFKFANPTDPCQLTEVHTHPDMDNIRLKYDK-----AGRVINIT 1199  
 Db 1727 DLYLVGYDADAGTWTAN-----ADQPHGR-KINKVSAGTTAAAGTYTYA 1771  
 Qy 1200 D-----NHGNTFETFDYTLGRL-----QNGQSVYGVDPANRVSQKITDLDDEL 1244  
 Db 1772 DGAMKQVREGSTTDTYTSRLRLATVTKTSGSGLTRTYDASGNLLVETTP----- 1825

Qy 1245 YRETMV---NEVRNGEMIRLRT-----GETIQAORA-----SKVLLTGTDSQOS 1289  
 Db 1826 --GETVASIGTBLRTDGSATATRYSGATTVAMRRTDGNNTVNGKITLMLGDTQAS 1883  
 Qy 1290 VILTSQKQ--NLQOAVSAYGKHS--TANDASILGNGEADPVSGYTHLGNRYSTDP 1346  
 Db 1884 TOIADVAAATGTAARRRRTTPGDERSGSLPTGYNHGFGKTEDNTYGLSL--GARADPS 1941  
 Qy 1347 LMFHPDPSLS-PFAGAGINPVYCGDPIINRSDPSGHLMOA-----ATGIGMGIA 1397  
 Db 1942 LGRFLSPDLSTPYDPONLSAISYSGNDPIINYDPSGLIKLNSDGTQCSGWMQKCGEYV 2001  
 Qy 1398 G-----TL-----LTATGMAIAAG-GIAA 1418  
 Db 2002 GGGGDTTPAPVQDQVLDLPREBNGWDADRLAQVWHYGOYTGCGGYWDAPVGDGDR 2061  
 Qy 1419 AIASTSTAL--APGALSVTSDITSIV-----SGALEDAPASISILG 1459  
 Db 2062 GMACFGRTACSEAFVWKETHDPAKARVATPCVENPKRCGADNGAYDSMKESAEPVI 2121  
 Qy 1460 VWSMGAGLAASAIKGTKLATHLGAFADGENALLK--STSESRIKMG-----V 1510  
 Db 2122 LLAGEMGA-----FSKVLKARCSEFKPTTRVLKDGKTKPLKIKRPGDLVEADP 2172  
 Qy 1511 TRSLDREI-----VRNEGQVIKDS---RGYTDNFMKGEOALIV 1548  
 Db 2173 TSGHREVRVTAVHLNHDDVLDLSIRGLDGRIGQLHTTARHRIMDTAQVMEQARLI 2232  
 Qy 1549 HGDKDGYLHYTEGNKH-----NKGPTHTPQVLDYLDKNNIVLDLTQGDGP-- 1557  
 Db 2233 TGKH-----VNTSGOHATITTSVLAORGADMDYLYVGLHTY-----YVLAGEPVL 2280  
 Qy 1598 VHLSCYKGS-----GAADKMAKYINRPVIAVSNKPTISQGLARIERKDFLKSTYHS 1651  
 Db 2281 VHGSCWSSNTRKTSVKNAGHMKH-----KSEFPNLMNAKEYIAGTDFLRST--- 2330  
 Qy 1652 YDPKILIGRTEKTVKXETPRP 1673  
 Db 2331 -DPS-VLTRTRANGDVIRFNP 2349

RESULT 7  
 US-10-282-122A-46565  
 : Sequence 46565, Application US/10282122A  
 : Publication No. US20040029129A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Wang, Liangsu  
 : APPLICANT: Zamudio, Carlos  
 : APPLICANT: Malone, Cheryl  
 : APPLICANT: Haebebeck, Robert  
 : APPLICANT: Ohlsen, Karl  
 : APPLICANT: Zyekind, Judith  
 : APPLICANT: Wall, Daniel  
 : APPLICANT: Trawick, John  
 : APPLICANT: Carr, Grant  
 : APPLICANT: Yamamoto, Robert  
 : APPLICANT: Forsyth, R.  
 : TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
 : FILE REFERENCE: ELITRA.034A  
 : CURRENT APPLICATION NUMBER: US/10/282.122A  
 : PRIOR FILING DATE: 2003-02-20  
 : PRIOR APPLICATION NUMBER: 60/191,078  
 : PRIOR FILING DATE: 2000-03-21  
 : PRIOR APPLICATION NUMBER: 60/206,848  
 : PRIOR FILING DATE: 2000-05-23  
 : PRIOR APPLICATION NUMBER: 60/207,727  
 : PRIOR FILING DATE: 2000-05-26  
 : PRIOR APPLICATION NUMBER: 60/230,335  
 : PRIOR FILING DATE: 2000-09-06  
 : PRIOR APPLICATION NUMBER: 60/230,347  
 : PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 46565  
LENGTH: 2234  
TYPE: PRT  
ORGANISM: Bacillus anthracis  
US-10-282-122A-46565

Query Match 4.3%; Score 386; DB 4; Length 2234;  
Best Local Similarity 18.7%; Pred. No. 5.7e-17;  
Matches 333; Conservative 236; Mismatches 633; Indels 608; Gaps 76;

21 FTGANNFTSAVSGGVDR-----TGLYNIQTLGHIIVG-----NGNL-----57  
748 YSKVGHASEISEHFEVRFSGDGGFICMPDYMASIPVLNGKVNATNGNFIEMSEKDITLSG 807  
58 -GPTLPITLISPLAKTDIGFGIGFNGLSV---YDRKSLISLS-----TGEN-102  
808 RGDVSVERTYNSQSKVGLFGTGWSSGLEBVRWADNGNLLISTDGANITFRTGDNK 867  
103 -----YKVIETDKTYKVLQKKLDNIRFEKDKLKNCTYRIHHSGLIEVL 145  
868 YQAPFTGYLEIKOVSGSEIKDKQIV-----TFYASGDQGR 905  
146 TGRNNAFDLKV-----KLLNPAGHAIYIDNFEATOPRLNRIYDDLGDHDIPL 196  
906 IETKOKYGTTTYEYDGAIRLSKVNASKELVL--QYDGNKKARV-----GPNKKT 959  
197 LMEYQGLITITLPPGQ--KSGYRTFLRLNQLNSINPISLGNENPLTWSFGY---250  
960 ITENYGDLLVSSITPBGKYKGYD-----NGVLTSIYDPOHTDAKPKTSYAYENDR 1013  
251 -----PIGKNGIL-----GQWITSMTAPGLKELTVNVSNNQGHHPQSANLPLPVYTL 301  
1014 LVKVTBPLGKATTLATYNTSGKEVTLINPKGRKTVYTTND-----AGNPLV-----1057  
302 MKQVPGAGPAPIDAEYSYTSNHYVG-----GNGIWMNKLDMLYGLMTEYN- 348  
1058 -KTVEDVGRNLITTSVEYNNANLVKTTTPKNGQETELATYDNGNVTSTVDEMGEKPEYMK 1116  
349 -YSTSRRKRYKE-----GHDQVIERITNTNNHLLTSECKQONGYIQTETAYYAI 400  
1117 DNDIITKATDDEDRKTTVAVGANTVEVSQTDGAN-----TSSVYIHDQYGNPIETSEKLS 1171  
401 IGHNPDSQSPQPLPKTKETWRSADNSVSEITETTFDSSGNPLTVIYDKTKQKILSP 460  
1172 AGGNLIONPS-FEM--NGTEKWKVVDNNSGSIS-----KDATPAPGGLSGESSLKITTK 1223  
461 STH--MEYYPAGEVDNCPPEPYGFTFRVKIIL-----QTPYDSEFKDP 503  
1224 ATNNDMGYIAAIEVLTLEPVTYTLTSGMYKTDLVNGAAPPVOSLINENAGIDGWHDIR 1283  
504 EKTIQYRISLIGSQSHVTLKIBERHYSATQLANSTLFQVYTTDKSELGRL-LKQTECTKGE 562  
1284 HNKVQOSTDMVNNQ--VTFKTEQ-----TRKVIYIQLVENGSGATSGSAWFDIOLKGR 1337  
563 NGKTYGVVHKFTYTKQ--DDTLQO-SHSITTH-----DNFTIHSQVRSRYTGLTFS 611  
1338 VSSSRPVLANSFEBENPDPGFVQWVRSCQCHERNDVSDSTGHSISYMER-----S 1391  
612 DTDTKDIVTQMSYDKLGRLLTRTLNSGTPYANTLLTYDYELANLQDDNRPPFVITTTDVG 671

1392 EYGPNDI-----GYRRVILNKKKATVTLTMSKSENVND-----1428  
672 NQLRNEFDAGRHVSQCLKDSGDGKFTYTHIQDYDQGHHTSTSYDYLTNGRQOTDP- 730  
1429 -----APDKLSKDYAVLAETYYODG-----TVNNYTSFSPSTNDW 1464  
731 -----DKVHLSKSKSYDMWQGLANTHWSGVSEKITYDPIITLPAKQLOSNS 777  
1465 NRSNAYIPAKPKIQKTEIFLFRKNKG-----KWFDDIRLBNALIKNE 1511  
778 NNVTGKEVTTYTPSQOPIQITTFDEAGHLQSCHTLTRDGMDBVRKSTDAIGQCTIYQYD 837  
1512 YD-NDGNVAVTY-----DEEGCK--NTFTYDASGNKKSSTDEKNTKLYDYN 1555  
838 NNRVVIQITLPDGTIVNRKAPSTDTLTDIVNNGISLQOQTFDGLSLRLOSODGGRW 897  
1556 KDMILTKVTLKNSTSVRYD-----1576  
898 AYTYSAGNDQCPSTVITPDSQFIHQYQPELDVAVLQVANSNETTQOFSPNPTGALLKAV 957  
1577 -----HNGNTTKSVMFQGKTQTHKTEYDVNKNVTYIDALNRRIEN-TYDENANKIKTKM 1631  
958 ABQSILTPYYPGRLKMNENINDMKKSYLMTLRLGSENGYTDLTGTLQKISRDPTHGRVQ 1017  
1632 PNGSILLESYVDTADRVGK-----RNGKDSFT-----FERDQNGVTK 1670  
1018 IKO--SISIKTLYAYDLNRHISQVTDLATGHLTTTVEPDGLNREIGRKLCDSSGHTLD 1075  
1671 VKDLVNGVERTKYTDVADR-----VTS-AT-----DSRGKID 1702  
1076 IQQSWL-----KTQGLANRIYKLVGLVQRTROYSDSNRNLNO--YKCDGACPTDXY 1126  
1703 -----MAYHDKANSKTEKAEQIVTQGY--TNKVSYD-YNTLDQNIKVTDSSQ-----1748  
1127 GHSIVTQNFYTDYGNITACTTTFADGTEDHATFK--ANP-----TDPCOLTEVYHT 1177  
1749 -----TYRFDYDQGNV-----RTYTAGNSGSGTFNPDQANKIDLVGINSLSLSRYE 1799  
1178 HPMPDNIRLKVDKAG-----RVINITDNHGTENFTYDTLGR-----LQ 1217  
1800 YDQSGNRRTYKHEGAGAGKVTETNFVYDPINQILNEVLNPGTYSYTDGFRNRTSVKIE 1859  
1218 NGQGS--YGYDPLRLVYSQKTDLDCEL-----1244  
1860 NGKETSIAATEFBGQVLVFGNESLTYDVNGNRSTDGKTKYTWNEDDQIVAITKQGENN 1919  
1245 -----YRETMVLVNEVRNGEMIRLRTGETI-----IAOQ 1274  
1920 AFATYKYDDEBNRIEKNVNGQVTRYFYDDDSINPLVETDNGTVLNRQYVVSADGARLAKM 1979  
1275 RASKVLLTGTDSQOQSYI--LTSDKONLSQEAAYSAYK--HKSTANDASILGNGERA 1327  
1980 AQCQTLTYHYNPRGADVAMTNOQKEVAVATEYDAMGNVLTSDTKGIAAD-NPFGYAGYMY 2038  
1328 DPVSGYTHLGNVGRSADPLMLREHPD-----SLSPFAG--GINPYSYCLGSPINR 1377  
2039 DKEIGMYTL-----IARYNEBHVFLSVDDPDEDEDDPVTMNGTYVADNNPVM 2088  
1378 SDPSGHLISQWAMTIGMGIAGLLTLATGGMALIAAGSIAAIASTSTYTLAAGALSPTS 1437  
2089 TDPDGMAM-----LVPVVAGAMVLAARFAGKAYAL-----RYGAKTGKKAIVKS 2131  
1438 DITSIVSGALDEASPRASSILGNVSNMGAGLAESALIKGYTLATHLGAFADGENALL 1497  
2132 -----GWDY-----GKVVAKSGMKWKGKSIACKIPRIHVGR--I 2163  
1498 KSTSESSRLKMGVTRSLDBREIVANNEGOYILKQHSRYTNPFMKGQALILVHDXGFLY 1557  
2164 KQDNDGKGYGVYISYTK-----KTGKRTYS-----SFEFHTPHNGKGY 2203  
1558 HTEGNKNG--KGPYTR 1572  
2204 HLQKNKYSKYQGMNR 2219



```

; APPLICANT: Yamamoto, Robert
; APPLICANT: Foreych, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282.122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 76214
; LENGTH: 843
; TYPE: PRF
; ORGANISM: Salmonella typhi
; US-10-282-122A-76214

Query Match      4.0%; Score 353.5; DB 4; Length 843;
Beat Local Similarity 22.8%; Pred. No. 2.3e-15;
Matches 223; Conservative 121; Mismatches 308; Indels 327; Gaps 48;

QY 542 YNTDKSELGR---LTKQTECKENGKTYSVHAKFTYTKODDTLQSHSITTHDNFTIH 597
DB 32 HHTDSGQYRLDMWMLAERSLCLVDSMGRT-----RCHMWDAGGLVAY 74
QY 598 RSQVRSRYTGR-----LPSDTPTKXIVTOMSDKXGRLLTRTNGSTPANTLTYYE 650
DB 75 RDEAGQMTTFRWSDEBRLLGNTDAAGGKMYVYDRLGH-----TETHD-P 120
QY 651 LNNLODNRPPF---VITTTDVNGNOLRNEFD-----GAGRHSQCL 689
DB 121 LGRVEQOMHFWHQPTETVDAAGVAMRYEYDERGNLQANSDDLKQTVYGYDHH-GQVV 179
QY 690 K--DSDDGKEYTHTHQYDEQG--RH-----HTSYSDYLTNCRQOTDPDKVHLSMS 738
DB 180 RITDARGGDKYL-----QWNEDEGLMHTDCSGQTAMFYDERRLRLRYDABE--NSTR 232
QY 739 KSYVNMQCIANTHMSYSEKITYDPTTLTATKOLQSNNSNVQCKEYTTTPEQCPIC 798
DB 233 YSYVGNHLETFEWMFADGTERYQPD-----AAGRIVKXTSPAG--QI 272
QY 799 TLFPDAGHLSCHTLTDGMDRVKRETDALIGCTIYOYDNNRYIOTLPGSTIVNRKYA 858
DB 273 TRWQ-----RDGGRVARKQTDATKRTAYEDAYGRLLTILNENGSYRFRY- 319
QY 859 PFSFTDLITDIRVNGISLGQOTPDGLSRLTGSOD--GGRVMAVYTSAGNDQCP----- 909
DB 320 -----DVLDRVTEQTPGGSBRAYGVNALNAVAVAYGGERG 356
QY 910 -----STVITPDGQFIHQYQPELDDAVLY-----ASREITQCP 944
DB 357 GEIRHGLERDAAGRLTAKITPE---TRREYVYDADRLLEIRRRHDAEGGEDEV- RF 412
QY 945 SYNFVTGALLKAVAG-----QSLTFPIYPSGRUKMENINDMKMSYLTMLRGLN 995

```

```

DB 413 SYSAGNLSSEFRAQVYLQHRVDVQGNRTETQMPDGR-----TLRLXY 456
QY 996 GYTDLTCTTICKISBTDHGRVLTQKDSIKTTLAYDLNRIHSGQVTDLATGMLTTTVER 1055
DB 457 G-----SGHLQOI-----NIGRVDVISHFT----- 476
QY 1056 DGLNREIGRLDCSSGHTLIDQSWLKTQOLANRIV--KUNGVLQR-----EYYSYSRNR 1110
DB 477 DHLHREVQR---SQGN-LMRRRYDRTGRLTRKLTCKMRGVVPEFTIDREIAYSGQDE 531
QY 1111 LNOYKCDGACPEPDKYSHSVTONFTYDIIGNITAC-HTTFADGETEDHATFKFANPTDPC 1169
DB 532 LK-----KRHSRQGVTDYF-YDTTGRITACGNEAYLDSWQDAA---ANLDDR 577
QY 1170 Q-----LTFVHHHTPMDPNIRLTKDKAGRVINITDNGNTENFTYDTL 1213
DB 578 QGETAQAAGSVVPEFNNITSYRGLH-----YRYEYGRVVEKRGANG-TQHYRWDAA 628
QY 1214 GRL-----ONGGASVYG--YDPLNRLVSQKTDLTDCELYYRETMVNVBRNGEMTLR 1265
DB 629 HRLTEVAVTGGRVRRYGVYDAPGRREVERHELDABEKFPNRTFTLMDGKRLAQECLGR 688
QY 1266 TGETIQAQR-----ASKYLLGTQD-SQOSVILTSQKMLSQEA-YSAV 1307
DB 689 SSSLIYSDRSGHPLARVDRAAPGEADRYLHYHTDVNGAPBEEMTDGCGNIWAGQVW 748
QY 1308 GK--HKSTANDASILGNGERADPVSGVTHLNGVRSYDPTLMRFPTDLSLDFG-AGG 1363
DB 749 GNLTHEKETRPVQGNLRFQCGYLDREFGHY--NLVYFPYDIDGKFTSGD---PIGLAGG 803
QY 1364 INPYSYLGDPINRSPSG 1382
DB 804 INLYQYA-PNPLSYIDPLG 821

RESULT 10
US-10-282-122A-47600
; Sequence 47600, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Maloney, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyckind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Foreych, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282.122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22

```

```

; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 47600
; LENGTH: 1515
; TYPE: PRT
; ORGANISM: Burkholderia cepacia
; US-10-282-122A-47600

Query Match      3.9%; Score 350; DB 4; Length 1515;
Best Local Similarity 22.1%; Pred. No. 1e-14;
Matches 232; Conservative 126; Mismatches 314; Indels 380; Gaps 59;

```

```

Qy 481 YGFTFRVYKIIQTFRYDSEFKDPEKFIQRYSLIGSQSHVTLKIEBHYSATOLLNSTLP 540
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 543 FQFGQFYAYVDEHGMWMTQMDTDQTVRYRDTAG-----RVVE-----TGTBQ 586
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 541 QYNTDKSELKLLKQTECTKENGKITYSVAKFTYTKODDTLQOSH-----ITHDNFTI 596
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 587 GYHT-----GRFIVEAGCT-----RVLDVDEMTYAVNDEGLVTAETDPLGHCYSEWEL 636
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 597 HRSQVR-----SRY--TGRLEFS-----DTDKDLYT-----QMSY 624
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 637 GRMARIDPLRRDXYRDERGQLTSSVSSGRVDPYDEQGLTGARLNGGSTITLEY 696
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 625 DKLRLLTRLTNSGTPYANTLYDY-----EL-----NNLDNRPPIVITTTDV-- 669
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 697 DHLRLIART-----EPDNKTTYRGPRGELLRVVQGRFRLDYDR-----LRLTDEL 748
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 670 -NGQQLRHEPFGARHNSQCLKSDGCKFTTHTQVDEGRHHTSYSDYLNGR--- 725
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 749 PTGARFRKIDALR-----LLEETSPDG-----HVTRYD-----YADGPNRGL 790
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 726 -QOTDPDKVHLMSKSYDNMGQIANTHWSYGVSEKITV---DPITLTATKOLQSNVQ 781
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 791 SAVTRPD-----GSVSARKN---SESLPFWMDPLRTIGR----- 824
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 782 TGEKVTYTPSQOPIQITLFDPAAGH-----LOSCHTLTRDGMDRVYK 823
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 825 -----TYGFPD--LITASIDAAGHATRFEDYHATRLTKVINALGETFYRYDAAGRLAA 876
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 824 ETDALGCTTIQYNNRVNIQTLBDCIIVRKAPSTFDLIDIRANGISLGG--QTF 881
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 877 EIDWGGATERYDRAVGRLLTKTLPDG-----QQWRITY 910
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 882 DGLSRLTQSGQGRVWATYTSAGNDQCPSTVITPDGFIHYQYQBELDAVLQVANSNEIT 941
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 911 DASRLIEIDAGDVLAIRYDASG-----RLASAVQGEHHTVT 949
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 942 QQFSYNPVTGALLKAAVAGQSITPIYPSGRMKMENINMKMSYLTWLRLENGYDTLT 1001
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 950 -RFAVDR--NGRLIGEDQGEELLRHYYDADGQRL--RMPPRRETYAV----- 993
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1002 GTICKISDTHGRVQIKDSIKTTLANDLNRIHGSQVTL-----ANG---HMLT 1050
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 994 -----DVSQALTVQV---QLTIRDGIGREIGRAGRPVAAQQYDALGRIRROLAG 1041
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1051 TTVEFGINREIGRLCDSSGHTLIDIOSWLTKOOLANRIYVLANGVLORTQOYSYDSNR 1110
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1042 PAVAFDALQADPARAL-----EQLTRQV-----YHDAAGQ 1072
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1111 LNOYKCDACEPT--DKYGHSTV-----TONFTYIYGNITACHTTPADGEDHATP 1160
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1073 LERVVD--TGADTLTYORDRGQITCAESLLQPSHFHYAVVMMI--AAHQORA----- 1121
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1161 KPAANTDDCOLTEVYHHT-----EDMPDNIIRLKVDAGAGVINT--DNHG---NTENFTYDT 1212
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1122 ----EVD-----AAHYRRGGLPEQVGYARVYDARGHTIEXTVQPGVRRPTQWYWDG 1171
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Qy 1213 LGRLE-----QNGQGSVYGYDPLNRVLSQKTDYLDCELYYRET-----MLVN---EV 1255
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1172 LNLVAVVYPERGVMA--YRDAFNRLREKQ-----QVGRFETVFKLMDGPMLAERVI EQ 1224
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1256 RNGEMIRLRTGTETI-----IAQGRASV--LITGTSQGSVILITSKONISQE 1302
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1225 RDG-----TYGGVNVWHLEPFGSFLPLAOTDGLPELITDQIGRPCTVDEQGRPWKA 1278
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1303 AVSAYGK--HKSTADA-----SILGNGERADPVSGVTHLGNYSYDPTLMRFH 1351
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1279 AYSLMGKLPVKRPPADADGATSIDTTLRFSGQMADEGTGLVNNLRY--YDPDSQYL 1336
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1352 TPDLSLSPFG--AGGINPYCYLADPINKSPDSG 1382
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1337 SAD---PIGLLGARQAY--VHDPQWIDPLG 1364
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 11
US-10-282-122A-43059
; Sequence 43059, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zykkind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H
; TITLE OR INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: EUTPA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 43059
; LENGTH: 1397
; TYPE: PRT
; ORGANISM: Escherichia coli
; US-10-282-122A-43059

Query Match      3.9%; Score 342; DB 4; Length 1397;
Best Local Similarity 21.4%; Pred. No. 3.3e-14;
Matches 256; Conservative 157; Mismatches 382; Indels 404; Gaps 63;

```



```

Db 218 RTQTFHREAAEFGSGITGVLT-DGAGRHFLVLTQOARAEARQAISGTEBSAPDPT 276
Qy 462 -----THWEYYPAGEVNCPEP-----YGF-----RFVK 488
Db 277 LPGTEYGRDNGIRLSAVMLTHDPEY-----ENLPAAPLVRYGWTFRGELAVYDSSGK 331
Qy 489 KIITQPYDSEFK-----DDEKFIQYRYSLSIGSHTVLKIERHYSATQLNSTL 539
Db 332 QVRSFTYDDKXRGMAVHRTGRPE--ICRYD--SDGRTEBQNLNAGLSYT----- 379
Qy 540 FQYNTDSELGRLKQTEC--TKENGKTVSVHKEFTYTKODDTLQOSSHITTHDNFTIH 597
Db 380 YQEKDRIITLDSLRREBVLTHQEGG-LKRVK-----EHADSQV 421
Qy 598 RSQVRSYTGRLFSDTQKQIVTQMSYDKGRLTRTLNSGTPYANTLTYDELANNOVD 657
Db 422 QSQDPA--VGRKQOTDAGRTTYSPPVYGLTRIT--TPGGRASAFY----- 468
Qy 658 NRPEVITTTDVNGNQLRNEFDGAGRHVSQCLKDS-----GDGKFTYIHTQYDEQGRH 712
Db 469 NHHSQLSAGTSPDGLAIRREYDEMGRLIOETAPDGDITRYKYNPHDLPATDATGSR 528
Qy 713 HTSTYSOYLNGROQTPDKVHLSMSKSYDNMGQIANTHMSYSEKITVDPI-TLTATK 771
Db 529 KTWMSRY--GQLLSFTDCSGYTRYDHRFGQVTAVHREBGLSQRAYDSRQGLIAYK 585
Qy 772 QLOSNANNOVGKVTYTPSQPIQITLPRDACHLOSCHLTPDGMGRKEDATGQ- 830
Db 586 DTQCHETRYE-----YNAADLTT--VIADPG--SRNGTOYDANGKA 623
Qy 831 -CTI-----YQYDNVNRVQITLPGDTIYNRKAPFSTDTLITDRVANGISLGQOTF- 881
Db 624 ICTTGGGLTSMEDAGRVIRLTSENGS-----HTFRVYDRLLOETGFPDRTQRYH 678
Qy 882 -DGLSRLOSODGRVAVYYSAGNDQCESTVITPDGFIHYQYQPELDAVLQVANSNEI 940
Db 679 HDLGLKIRSEDEGLV-----THMYD-BADRLTHRTVANGET 714
Qy 941 TQGSYVPVYCAL--LKAVALGOSLTPY-YPS-GRUKMEI-----NDM--KMSY 986
Db 715 AERQOYB-RGWLTDHSHISGHRVYHYGDSKGRASELTVHNPOTWELLMOHETRH 773
Qy 987 LMTLRLNGEYTDLTGTIQRISDTHGRVQIKOSIKITLMTYDNLNRHIGSOVTDATG 1046
Db 774 AYNAQGLAN--RCIPDLSRAVEMLTYG-----SGMLSCKMG 808
Qy 1047 HMLTTYEF--DGLNREIGRGLCDSSGHTLDIOQSMKTOQLANRYKLVGLORTQOYS 1104
Db 809 D--TPLVEYTRDLRHRETLR-----SFGRYELTAYTPAQLOSG--HLSLSL-DRDYT 858
Qy 1105 YDSNRNLNOYKCDACEPTDKYGSIVYQNTYIYGNITACHTTPAD-----GTEDHA 1158
Db 859 WNDNGELIR-----ISSPRO-----TRSYSSTRLTGVHTTAANLDIRIFYTTPA 906
Qy 1159 TPKFANPTDPCQLTEVHHTHPD-----MPDN-----IRLKYDKAGRVINITD----- 1200
Db 907 GNRLPDP-----ELHPDLSLMPDPNRIRAKDAHYLYRYRGRGLTEKIDLIREGV 956
Qy 1201 ---NHGNTENFTYDGLRLONGOSVYG-----YDPLNRLVSOKTDLDEL----- 1244
Db 957 IRTDERTHRHYDSGRRLVHYTRQYABPLVESRYLYDPLGRVAVARWRERENDLGM 1016
Qy 1245 ---YY--RETMVNE-----VRNGEMIRLR-----TGETIIAQORA-- 1276
Db 1017 SLSRKQVYTWYMGDGLTITQNDRTRIQTIYQGSFTPLRVTATAGELAKTORSLAD 1076
Qy 1277 ---SKVLTGTSQOSVILTSDKOMLSOA----- 1303
Db 1077 TLOQSGEDGVSVPVLYQMLDRLESEILA--DRVSESRMLASCGILTVAQOMOSOM 1133
Qy 1304 ---YSAVGRKXSTANDAS-----ILQYNG 1324
Db 1134 DPVYTPARKIHLVYCHDRGLPLALISTEGTAYVAEYDEWGNLNLNENPHOLOQLIRLPG 1193

```

```

Qy 1325 ERADPVSGVTHLNGYRSYDPTLMRFTTPDLSPPFG-AGGINPYXCIGAPINRSDPSG 1382
Db 1194 QQIDBSGLVY--NRHRYDPLQGRYITTD---PIGLKGMNFIQYPL-NPISNIDPLG 1246

RESULT 12
US-10-282-43060
: Sequence 43060, Application US/10282122A
: Publication No. US20040029129A1
: GENERAL INFORMATION:
: APPLICANT: Wang, Liangsu
: APPLICANT: Zamudio, Carlos
: APPLICANT: Malone, Cheryl
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Karl
: APPLICANT: Zykkind, Judith
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John
: APPLICANT: Carr, Grant
: APPLICANT: Yamamoto, Robert
: APPLICANT: Forsyth, R.
: APPLICANT: Xu, H.
: TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
: FILE REFERENCE: ELITRA.034A
: CURRENT APPLICATION NUMBER: US/10/282,122A
: CURRENT FILING DATE: 2003-02-20
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/230,335
: PRIOR FILING DATE: 2000-09-06
: PRIOR APPLICATION NUMBER: 60/230,347
: PRIOR FILING DATE: 2000-09-09
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/267,636
: PRIOR FILING DATE: 2001-02-09
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: Remaining Prior Application data removed - See File Wrapper or PALM.
: NUMBER OF SEQ ID NOS: 78614
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 43060
: LENGTH: 1411
: TYPE: PRT
: ORGANISM: Escherichia coli
: US-10-282-122A-43060

Query Match 3.8%; Score 340.5; DB 4; Length 1411;
Best Local Similarity 20.4%; Pred. No. 4.3e-14;
Matches 287; Conservative 172; Mismatches 459; Indels 490; Gaps 70;

Qy 418 KTEFW-RSADNSRSTRSETTEPDSGNPLFKV-----KDKTKOKIS-----PS----- 461
Db 218 RTQTFHREAAEFGSGITGVLT-DGAGRHFLVLTQOARAEARQAISGTEBSAPDPT 276
Qy 462 -----THWEYYPAGEVNCPEP-----YGF-----RFVK 488
Db 277 LPGTEYGRDNGIRLSAVMLTHDPEY-----ENLPAAPLVRYGWTFRGELAVYDSSGK 331
Qy 489 KIITQPYDSEFK-----DDEKFIQYRYSLSIGSHTVLKIERHYSATQLNSTL 539
Db 332 QVRSFTYDDKXRGMAVHRTGRPE--IRYRD--SDGRVTEBQNLNAGLSYT----- 379
Qy 540 FQYNTDSELGRLKQTEC--TKENGKTVSVHKEFTYTKODDTLQOSSHITTHDNFTIH 597

```

Db 380 YQEKDRTITDSDLREVLHTQCEAG-LKRNVKK-----HEADSVT 421  
Qy 598 RSQVRSKRTGRLFSDDTKDVLVQMSYDKLGRLLTRTINSTGYPAANTLYTELANI 657  
Db 422 QSQPDA--VGRRAQOTAAAGRTTESPDVYGLTRIT--TPGASAFYNNHQ--- 473  
Qy 658 NRPFVITTTDVNNGQJANEFDPAGRHVSQCLXSD-----GDGKFYTIHTQYDEGRH 712  
Db 474 ----LTSATPDDLELRREIDELGRLIQETAPBDITRYVNDPHSLPCATDAGSR 528  
Qy 713 HTSTYSDYLTNGROQTPBDKVLHLSMSKSYDNWGOJANTHMSGVSEKITVDP--TLTATK 771  
Db 529 KMTWMSRY---GQLSFTDCSGVYTRVDHDFGQMTAVHREBGLSQYRAYDSRQGLAVK 585  
Qy 772 QLOGNSNNVQTKGKVTYTPSQRPQITLFEBAHQ-----SCHTLTDGMDRVKKE 824  
Db 586 DTQGHETRYE-----VNIAGDLTAVIAPDGSRNQTOYDAMGKAVRT 626  
Qy 825 TDAIGQCTIYQDYNVRYOITLPDGTIVNKRKAPFSTDTLITDIRVNGISLGQTF--D 882  
Db 627 TQG--GLTRSMETDAAGVIRLISENGS---HTFRYVDLDRLIQETGFPDGTORXNHD 680  
Qy 883 GLSRLTQSQDGRWAWYYSAGNDQCPSTVITPDQFIHYQOPELDAVLQVANSNETIQ 942  
Db 681 LTGLKLISEDEGLV-----THMYD--EADRLTHRYVKGETAE 716  
Qy 943 QPSTNPVTGAL--LKAVALGOSLPIYY--PSGRL--KJENINDMKMSYLM----- 988  
Db 717 RWOYDE--RGMVLTDSHISEGRHVA--VHYRYDEKGLTGEROTVHPQTEALLMOHETRHA 774  
Qy 989 -TLGLENGYTDLGTIOKISRDTGRVTOIKDSIKTLLAVYDLNRIIGSOVDTLATGH 1047  
Db 775 YNAQGLAN--RCIPDSIPAVEMLYG-----SCYLAGMKLGD----- 809  
Qy 1048 MLTTTVEF--DGLNREIGRKLCDSSGHTLDIQSSMLKTQOLANRIVKLVGLQSTEOYSY 1105  
Db 810 --TLPVEVTRRLRHRETLR-----SFGRYELTAVTPAGQOQ--HNLNLSL--DRDYTW 859  
Qy 1106 DSRRLNIOYKCDGACFPDCKYGHSLVTONFTYDIYGNITACHTTPADGTEHAPFKFANP 1165  
Db 860 NDNGELIR-----ISSPRO-----TRSYSTGTGRLTGVHTAA-----NLDIRIPYA 902  
Qy 1166 TDPC--OLTEVHTHPD-----MPDN-----IRLKYDKAGRYVNTD-----NH 1202  
Db 903 TDPGKNRPD--PELHPDSTLSMMPDNRIARDAHYLYRDRGRLEKTDLIPBGYTRTD 961  
Qy 1203 GNTENFTYDTLGRLONGO-----GSVYGYDPLNRLVSOQTDLLDCEU----- 1244  
Db 962 ERTIRHYHDSQHRLLVHYTRQYEEPLVBSRYLIDPLGRVAKRWRRERDLDGMMSLSRK 1021  
Qy 1245 ---Y---RETMVNE-----VANKEKIRLR---TGETTIAQORA----- 1276  
Db 1022 POUVWYMGDGRLLTIQMDRSRIQTIYQPSFTPLIRVETATGELAKTORRSIADALQOS 1081  
Qy 1277 -----SKVLLTGTDSOOSVLTSDKONLSQEA----- 1303  
Db 1082 GGBDGSVFPFVPLQMLDLRLESEILA--DRVSEESRWMLASCGLTVEQMKOMDPYTT 1138  
Qy 1304 -----YSAYGKHKSSTAND-----ASITLYNGERADP 1329  
Db 1139 PARKIHLVHCHRGRLPLALISTEGATAWCAEYDEWGNLNBENPHQLOQLRLPQOQYDE 1198  
Qy 1330 VSGVTHLNGKRSYDPTLMRFTPLSLSPFG--AGSINPYSYCLGPIRNSPDSGHLMSQA 1388  
Db 1199 ESGLYY--NRHRYDPLQGRYITQD---PIGLKGMNLYGYOL--NPISIDIPLGISMWED 1252  
Qy 1389 WTGIGMAGLILITATGMAIAAAGIAAALASTTALAAGALSYSDITSISYGALE 1448  
Db 1253 -----AKSG--ACTNGACGTLSA-----MIPDPKDSIDSTAY 1283  
Qy 1449 DASPAASSILGVNWSMGMAAGLABSALIKGTGLATHLGAFADGNNALLKSTSESSRIKW 1508

Db 1284 DALINKINS-----QSIDCKEFA----- 1301  
Qy 1509 GYVRSUDREIVRNEBGOVTKDSRGYTDNFMKGEO-----ALLVHGKDGFLYHTEG 1561  
Db 1302 -----GLICDNGSRIFSTAPNRGGRKGSYFPNPPCNGTEKYSAYHTHG 1346  
Qy 1562 NKHNGK--GPYTRHTPEOLVYLDKNNI 1587  
Db 1347 ADSHGEYWDIEIFSGKDEKIYK--SKDNNI 1373  
  
RESULT 13  
US-09-912-020-340  
Sequence 340, Application US/09912020  
Patent No. US20020045592A1  
GENERAL INFORMATION:  
APPLICANT: Zykend, Judith  
APPLICANT: Olsen, Karl L.  
APPLICANT: Forsyth, John  
APPLICANT: Froelich, Jamie M.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN  
FILE REFERENCE: ELITRA.001DV1  
CURRENT APPLICATION NUMBER: US/09/912.020  
CURRENT FILING DATE: 2001-07-23  
PRIOR APPLICATION NUMBER: 09/492,709  
PRIOR FILING DATE: 2000-01-27  
PRIOR APPLICATION NUMBER: 60/117,405  
PRIOR FILING DATE: 1999-01-27  
NUMBER OF SEQ ID NOS: 485  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 340  
LENGTH: 1426  
TYPE: PRM  
ORGANISM: B. Coli  
US-09-912-020-340  
  
Query Match 3.8%; Score 339; DB 3; Length 1426;  
Best Local Similarity 20.2%; Pred. No. 5.5e-14;  
Matches 257; Conservative 150; Mismatches 409; Indels 458; Gaps 60;  
  
Qy 572 KFTYTKDDTLQOSHSTTHDNFTIHSQVRSKRTGRLFSPTDKDVLVQMSYDKLGRLL 631  
Db 317 RYTYTEAGELL---AVYDRSNTQVRAFTYDAQHGMVAVHRYVAGRPDMRYDDTGRV 372  
Qy 632 TRTNSGTPYANTLYTYEELNLODDNRPPFVITTTD----- 668  
Db 373 EQLPAG-----LSYK---LYEQR---ITVDSLNRREVLHTEBGAGLGRVAKKE 418  
Qy 669 -VNGNOLRNEFDGAGRHSQCLXSDSGKFTYTIHTQOYD-----EGRHHTSYSDYL 721  
Db 419 LADSVYRSRGVDAAGRLTAQ--TDAAGRTEYGLNVVSGDITDITPDGRETKEFYND-- 474  
Qy 722 TNGEQC---DPRVHLHLSMSKSYDNWGOJANTHMSYGVSEKITVDTPLTATKLOGNSN 778  
Db 475 --GNQLTAVVSPD--GLSREXYDEPRLV-----SE 502  
Qy 779 NVQNGKEVT--TYPSQRPQITLFEBAHQSCHTLTRDGMVRYKETAIGQCTIYQYD 837  
Db 503 TNSGSEVRYRYDDAHSELPTTTDANG--STRQMTWSKYGQLAFTDCGYOTRIEYD 559  
Qy 838 NYNNEVIOITLPDGTIVNKRKAPFSTDTLITDIRVNGISLGQTFDGLSRLTQSDG--GRV 896  
Db 560 RFGM-----TAVHRE-----EGSL--YRAYDNGRLTSVXDAQGRE 595  
Qy 897 WATYSGANDQCPSTVITPDQFIHYQO----- 925  
Db 596 TRYRYNAGDIL--TAVITPGRNRSQYDAMGKAVSTTQGLTRSMEXDAAGVLSLTNE 653

```

Qy 926 -----BELDAVLQVANSNEITQOFSYNPVTGALLKAVAGOSLTPITYPSGRKME 976
Db 654 NGSHSVSYDALRLVQGGGFDGRTQRYHD-LTGKLTQSEDEGLVLTWYDESDRITHR 712
Qy 977 NINDMKKSYLWTLRGLNGEYDTLGTIOKISBDTHGRVTOIKDS-----IKTTLAYDDL 1032
Db 713 TVNGEPARQOY-----DGHGMLTDISHLSGSHVAVHYGDDK 751
Qy 1033 NRHIG--SOYTDLATGML----- 1049
Db 752 GRLTGECQVTEPNPTEGELLWQHETKAVYNOGLANRVTPDLSPEVWLYTSGYLAKMKL 811
Qy 1050 --TTTVAF--DGLNREIGRKLCDSSGH--TLDIOQSWLTKOQLANR----- 1089
Db 812 GCPPLVYTRDLRHLRETVRSGMAGSNAVELSTTPACQLOSGLNSLVYDRDYGWS 871
Qy 1090 ----YVLNGVLQRTQOYSYDSRNINQYKCDGAC-----PTDKYGHSI----- 1130
Db 872 DNGDLVHISGPRQ-TREYGSATGRLESVRLAPDLDIRIPYATDPAGNRLPDELPDS 930
Qy 1131 -----VTQN-----FTYDIYGNITACHTTPADG--TEDNATFKFANPTDPCQLTEV 1174
Db 931 TLTVPMNRKLAEDAHVYRDEYGRLEKTDRLPAGVIRTDDEBT----- 975
Qy 1175 HHTHPDMDNRLKYDKAGRVINITD-NHGN--TENFTYDTLGR----- 1215
Db 976 HHNH-----YDSQHLVFYTRIQHGEPLVESKYLDPDLGRRAKAVMRERDLTG 1025
Qy 1216 ----LQNGQSVYGYDPLNRLVSOQTDLDCELYR----- 1247
Db 1026 WMSLSRKEVTVYGMWG-DRLTIVQTDTRIQTVYEGSGFTPLIRVETENGEREKQRRS 1084
Qy 1248 --ETMLVNEVRNG-----EMIRLL-RTGETTIA-----OQRAKY 1279
Db 1085 LAETLQOEGSGENGHVVPAPALVRLDLRLEBIRADRVSSBRAMLAQCGLYEQLARQV 1144
Qy 1280 LLAGTDSQOS-----VILTSDKONLSQEA-VSAYGKHKSTANDASIIIGYN----- 1323
Db 1145 EPEETPRARKAHLHCHDRGLPLALISDGTAMSAEIDEMGNQINENPHNV-YQPYRL 1202
Qy 1324 -GERADPVSQVTLGNGYSYDPTLMRFHPDLSPPG-AGGINPVSYCLADPINSRPS 1381
Db 1203 PGQOHDEBSGLY--NHRHYDPLQGRYITQD--FMGLKGWMLYQYPL-NPLQOQDPM 1256
Qy 1382 GHLSQWMTGIGAGIAGLLTATGMAIAAGIAAIAIATSTTALAFGLASTSDITS 1441
Db 1257 GLL--QTMDDARSG-----ACTGV--CGVLSRIIGPSKFDSTADALD----- 1296
Qy 1442 IVSGALDASPKASSILGWYMGWGAAGLASAIKGTAKLATHLGAFADGEMNALKSTS 1501
Db 1297 ----ALKEQNRS-----LQNDMEYSIGVCKDTNG-----KYFASKAETDNLK--K 1336
Qy 1502 ESSRIKMGVTRSLDREIVRNEBGOVIXDHSRGYTDNFMGKGBQAILVHGDK--DGLYHT 1559
Db 1337 ESYPLKKCKPTGTRVAAHYTHG--ADSHGDYVDEFFSSDKVLVSKONNLEAFYLAT 1393
Qy 1560 EGNKH--NGKPY 1570
Db 1394 PDGRFEALNNKGEY 1407

```

## RESULT 14

```

US-10-282-122A-42617
Sequence 42617, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haeelbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zykkind, Judith
APPLICANT: Wall, Daniel

```

```

APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA 034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remainder Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 42617
LENGTH: 1426
TYPE: PR1
ORGANISM: Escherichia coli
US-10-282-122A-42617
Query Match 3.8%; Score 339; DB 4; Length 1426;
Best Local Similarity 20.2%; Pred. No. 5,5e-14;
Matches 257; Conservative 150; Mismatches 409; Indels 458; Gaps 60;
Qy 572 KPTTKQDDTLQOSSHITTHDNFTIHSQVRSRYGRLPESDVTQKQIVTQMSYDKGRLL 631
Db 317 RYTYTEAGELL--AYDRSNTQVRAFTYDAQHPGMAVHARYAGRPENKRYRDTGRV 372
Qy 632 TRTLNAGTPYANTLTVDYELANLQDNRPEFVITTD----- 668
Db 373 EQLNPRG-----LSRY--LYEQR-----ITVDSLRRREVLTBEGAGLKVYKKE 418
Qy 669 -VNGQNRNEFDGAGRHVSQCLKDSQDGKFTYTHQYD-----EQGHHTSTSDYL 721
Db 419 IAGGSVTRSGYDAAGRLTQ--TDAAGRRTYGLANVSGDITDITPDGRTKRYND-- 474
Qy 722 TNGRQOT--DPPKHLMSKSYDNMGQANTMTMSYGVSEKTIYDPTTLATKQLQNSN 778
Db 475 --GNQLTAVVSP--GLSRRREYDEPERLV-----SE 502
Qy 779 NVQTKREV-TYTPSOQPIQITLFDAGHLQSCHTLTRDQMDVRKETAIGQCTIYQYD 837
Db 503 TSSGFTVRVRYDDASELPATTTDATG--STRQMTWSRYGQLAFITDQSGVQTRYED 559
Qy 838 NNRKVIQITLPDQTVNRKAPSTDTLTDIRVNGISLQOQTFDGLSRLOSQDG-GRV 896
Db 560 RFQOM-----TAVHR-----EGISL-YREYDNRGRLLTSYKDAQRE 595
Qy 897 WATYSAGNDQCSYTPDQGFHYQY----- 925
Db 596 TRYEYNAAGDL-TAVITPDGNSRSETOYDAMGCAVSTQOGLTRSMETDAAGVISTLNE 653
Qy 926 -----BELDAVLQVANSNEITQOFSYNPVTGALLKAVAGOSLTPITYPSGRKME 976
Db 654 NGSHSVSYDALRLVQGGGFDGRTQRYHD-LTGKLTQSEDEGLVLTWYDESDRITHR 712

```

```

Qy 977 NINDMKKMSYLMTLRGLNGTYD.LGTIIOKISBPTHGRVYOIKDSS-----IKTLNVDL 1032
Db 713 TVNGEPBAQWOY-----DGHGMLTIDISHLSGHRVAHYGYDDK 751
Qy 1033 NRHIG--SOVYDLATGML-----1049
Db 752 GRLTGBEOQYBENPRTGELMQHETKHAVNEOGLANRVTPDLS.PPVEWLTGSGYLAKML 811
Qy 1050 --TTTVEF--DGLNREIGRLCDSSGH--TLDIOQSWLKTQOLNR-----1089
Db 812 GGTPLVEYTRDLRLRETVRSFGMAAGNAAYELSTYTPAQLOSOHLNSLVYDRDYGWS 871
Qy 1090 ----IVKLVGLQRTBOYSYDSRNLNOYKCDGAE-----PTDKGHSI-----1130
Db 872 DNGDLVRISGPRQ--TREYGSATGRLESVRLAPDLDIRIYATDPAGNRL.PDELIHDS 930
Qy 1131 -----VTQN-----FTYDIYGNITACHTTFADG--TEDHATFKFANPTDPCQLTEV 1174
Db 931 TLTVWPNRILAEADAHVYRHDEYGRLTTEKTDRI.PAGVIRTDERT-----975
Qy 1175 HHTHPMDNIRLKYDKAGRVINITD--NHGN--TENFTYDTLGR-----1215
Db 976 HHYH-----YDSQHRLVFYTRI.QHGEPVLESRYLXPLGRMAKRVWRERBDLTG 1025
Qy 1216 ----LONQGSVGYDPLNRLVSOIKDTLDCELYR-----1247
Db 1026 WMSLSRKEPVTWYGMG--DRLTIVQDTITRI.QTYEPGSPFLIRVETENGEREKAQRS 1084
Qy 1248 --ETMLVNEVNG-----EMIRL--RTGETTIA-----QORASKV 1279
Db 1085 LAETLQOEGSENGHVPAELVRLDLREIEIRADRVSSERAMLAQCGLTVEQLARQV 1144
Qy 1280 LLTGTDSQOS-----VILTSQKUNISOE--YSAVGHKHSSTANDASILGN---1323
Db 1145 EPEYTPAKKALHYCDHRGLPLALISEBDNTAMSEYDWMGNQENBEPHNV--YQPYRL 1202
Qy 1324 --GERADPVSGLTHGNGYRSYDPTLMRPHTPDLSLSPFG--AGGINPYSCLDGPINRSDPS 1381
Db 1203 PGQGHDESSGLY--NRHRYDPLQGRYITOD--PMSLKSGMMLYQVPL--NPLQOQIDPM 1256
Qy 1382 GHLSMOANTGIGMGIAGLLTITATGMAIAAAGIAAIASTTTALFALSVTSDITS 1441
Db 1257 GLL--QTWDDARSG-----ACTGCV--CGVLSRIIGPSKFDSTADALD-----1296
Qy 1442 IVSALBEPASPRASSILGWSMGAGLAASAIKGTAKLTHLGAPEDEENMLIKSTS 1501
Db 1297 ----ALKETQURS-----LCNDMEYSIGVCKDTNG-----KYFASKAETDNLK--K 1336
Qy 1502 ESSRIKMGVTRSLDEIVRNEBGVYIKDHSRGYTDNFMKGEOAILVHGDK--DQFLYHT 1559
Db 1337 ESYPLKRCPTGTDRVAAYHTHG--ADSHQDYDEFFSSSDKNLVRSKNNLEAFYIAT 1393
Qy 1560 EGNKH--NGKGPY 1570
Db 1394 PDGRPEALNNKGEY 1407

```

```

; PRIOR FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 340
; LENGTH: 1426
; TYPE: PRT
; ORGANISM: E. Coli
US-10-771-241-340

Query Match      3.84; Score 339; DB 5; Length 1426;
Best Local Similarity 20.24; Pred. No. 5.5e-14;
Matches 257; Conservative 150; Mismatches 409; Indels 458; Gaps 60;

Qy 572 KFTYTKODDTLQOSSHITTDNFTIHRQVRSRYTGLEFSDTDTKDIVTQMSYDKLGRLL 631
Db 317 RYTYTTEGELL--AVYDSNSTQVAFYVDAGHPGMAHRRVAGREMRERYDDTGRVY 372
Qy 632 TRTINSSTPYANTLYDYELNLLQODNRPPFVITTD-----668
Db 373 EQLNPAQ-----LSYRY-----LYEQDR-----ITVTDLSLRREYLTGEGAGLKRIVYKKE 418
Qy 669 --VNGNQLRNFPDAGRVVSGQLKDSGDGKFYIIHQVYD-----EGGRHHTSYSDYL 721
Db 419 LAQGSVTRSGYDAAGRLTAQ--TDAAGRTEYGLNVVSGDITDITPDGRTKRYND-- 474
Qy 722 TNGRQOT--DPRKVLHMSKSYDNMGQIANTHWSYGVSEKITVDPIITLTATKQLQGSNSN 778
Db 475 --GNQLTAVVSPD--GLSRSRYDEPGRVY-----SE 502
Qy 779 NVQTKKEVT--TYTPSQPQIQTLEDEAGHLQSCHTLTRDGMDRVKRETDALGOCTIYQYD 837
Db 503 TSSGSETVRYRDAHSELBATTTDATG--STROMTWSRYGQLAFTDCSGYQTRYEYD 559
Qy 838 NNRKRVIOITLPDGIYVRKXAPSTDTLIDIVNGISLQOQFPDGSRLTQSGDS--GRV 896
Db 560 RFGQM-----TAVRE-----EGISL--YRRYDNRGLTSVDAQGRE 595
Qy 897 WAAVYSAGNDQCESTVTPDQGFVHYQY-----925
Db 596 TRYEYNAAGDL--TAVITTPGNSSETOYDAMGRAVSTTQSGLTRSMEMYDAAGRVISLTNE 653
Qy 926 ----BELDVAVLQVANSBITQOFSYNPVYTGALLKAVAGQSILPTIYPSGRLMKE 976
Db 654 NGSHSVESYDALRVLVQGGFGDGRTRYHYD--LTGKLTQSEDEGLVLWYDESDRATTHR 712
Qy 977 NINDMKKMSYLMTLRGLNGTYD.LGTIIOKISBPTHGRVYOIKDSS-----IKTLNVDL 1032
Db 713 TVNGEPBAQWOY-----DGHGMLTIDISHLSGHRVAHYGYDDK 751
Qy 1033 NRHIG--SOVYDLATGML-----1049
Db 752 GRLTGBEOQYBENPRTGELMQHETKHAVNEOGLANRVTPDLS.PPVEWLTGSGYLAKML 811
Qy 1050 --TTTVEF--DGLNREIGRLCDSSGH--TLDIOQSWLKTQOLNR-----1089
Db 812 GGTPLVEYTRDLRLRETVRSFGMAAGNAAYELSTYTPAQLOSOHLNSLVYDRDYGWS 871
Qy 1090 ----IVKLVGLQRTBOYSYDSRNLNOYKCDGAE-----PTDKGHSI-----1130
Db 872 DNGDLVRISGPRQ--TREYGSATGRLESVRLAPDLDIRIYATDPAGNRL.PDELIHDS 930
Qy 1131 -----VTQN-----FTYDIYGNITACHTTFADG--TEDHATFKFANPTDPCQLTEV 1174
Db 931 TLTVWPNRILAEADAHVYRHDEYGRLTTEKTDRI.PAGVIRTDERT-----975
Qy 1175 HHTHPMDNIRLKYDKAGRVINITD--NHGN--TENFTYDTLGR-----1215
Db 976 HHYH-----YDSQHRLVFYTRI.QHGEPVLESRYLXPLGRMAKRVWRERBDLTG 1025
Qy 1216 ----LONQGSVGYDPLNRLVSOIKDTLDCELYR-----1247
Db 1026 WMSLSRKEPVTWYGMG--DRLTIVQDTITRI.QTYEPGSPFLIRVETENGEREKAQRS 1084

```

```

Qy 1248 --ETMLVNEVANG-----EMIRLL-RTGETTIA-----OORASKV 1279
Db 1085 LAETLOQEGSENGGVPPAEVLVLLDLREHEIRADRVSSSESRAMLAQCGLTVEQLAROV 1144
Qy 1280 LITGTSQOQ-----VILTSDKONLSQEA-YSAYGKHSTANDASILGYN----- 1323
Db 1145 EPEYTPARKAKHLVHCDHRGLPLALISEDGNTAMSAEYDEMGNQINENPHIV--YQPYRL 1202
Qy 1324 -GERADPVSGVTHLNGVRSYDPTLMRFHTPDSLSPFG-AGGINPYSCLDPINRSDPS 1381
Db 1203 PGQOHDEESGLY--NHRRTYDPLQGRYITQD---PMGLKGWMLYQYPL-NPLQOQIDPM 1256
Qy 1382 GHLSMQAMTGIGMGIAGLLLTATGMAIAAAGIAAIASTSTTALAFGALSVTSDITS 1441
Db 1257 GLL--QTMDDARSG-----ACTGV-----CGVLSRIIGPSKFDSTDAALD----- 1296
Qy 1442 IVSGALEDASPKASSILGWVSMGMAAGLAESAIGGTCLKLATHLGAPAEDEGNALLKSTS 1501
Db 1297 ----ALKETQNR-----LCNDMEYSGIVCKDTNG-----KVFASKAETDNLK--K 1336
Qy 1502 ESSRIKMGVTRSLDRREIVRNEEGVITQDSKGYTDNFMGKEQALIVHGDK--DGFLYHT 1559
Db 1337 ESYPLKRRKCPGTGTRVAAYHTHG---ADSHGDYVDEFFSSSDKXVLVRSKONNLEAFYLAT 1393
Qy 1560 EGNKH--NGKGPY 1570
Db 1394 PDGRFEALNNKGEY 1407

```

Search completed: January 30, 2006, 09:53:03  
 Job time : 115.198 secs

**THIS PAGE LEFT BLANK**

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: January 30, 2006, 09:46:19 ; Search time 13.3373 Seconds

(without alignments)  
1358.372 Million cell updates/sec

Title: US-09-889-874a-23

Perfect score: 8879  
Sequence: 1 VYIKFLKLPFRIMSDNNEF.....PKIILGRTEKYKPTFRP 1673

Scoring table: BLOSUM62

Gapop 10.0 , Gapept 0.5

Searched: 75621 seqs, 10829074 residues

Total number of hits satisfying chosen parameters: 75621

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

## Database :

1: Published Applications AA New:  
2: /cgn2\_6/ptodaca/1/pubpaa/US08\_NEW\_PUB pep: \*  
3: /cgn2\_6/ptodaca/1/pubpaa/US07\_NEW\_PUB pep: \*  
4: /cgn2\_6/ptodaca/1/pubpaa/US09\_NEW\_PUB pep: \*  
5: /cgn2\_6/ptodaca/1/pubpaa/US10\_NEW\_PUB pep: \*  
6: /cgn2\_6/ptodaca/1/pubpaa/US11\_NEW\_PUB pep: \*  
7: /cgn2\_6/ptodaca/1/pubpaa/US11\_NEW\_PUB pep: \*  
8: /cgn2\_6/ptodaca/1/pubpaa/US60\_NEW\_PUB pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	285	3.2	2721	7 US-11-096-051-10	Sequence 10, App1
2	285	3.2	2725	7 US-11-096-051-8	Sequence 9, App1
3	284.5	3.2	2715	7 US-11-113-424-51	Sequence 51, App1
4	283	3.2	2376	7 US-11-096-051-4	Sequence 4, App1
5	283	3.2	2715	7 US-11-096-051-2	Sequence 2, App1
6	273.5	3.1	1094	6 US-10-821-234-1097	Sequence 1097, App
7	272.5	3.1	2333	6 US-10-453-372-170	Sequence 170, App
8	272.5	3.1	2662	6 US-10-453-372-114	Sequence 114, App
9	272.5	3.1	2724	6 US-10-453-372-148	Sequence 148, App
10	272.5	3.1	2733	6 US-10-453-372-136	Sequence 136, App
11	272.5	3.1	2733	6 US-10-453-372-142	Sequence 142, App
12	272.5	3.1	2733	6 US-10-453-372-146	Sequence 146, App
13	272.5	3.1	2733	6 US-10-453-372-150	Sequence 150, App
14	272.5	3.1	2733	6 US-10-453-372-154	Sequence 154, App
15	272.5	3.1	2765	6 US-10-453-372-116	Sequence 116, App
16	271	3.1	2759	6 US-10-453-372-168	Sequence 168, App
17	266.5	3.0	2515	7 US-11-113-424-53	Sequence 53, App1
18	253.5	2.9	2769	7 US-11-113-424-54	Sequence 54, App1
19	253	2.8	2769	7 US-11-113-424-14	Sequence 14, App1
20	226	2.5	2725	7 US-11-113-424-52	Sequence 52, App1
21	216	2.4	2053	7 US-11-013-759-9	Sequence 9, App1
22	203.5	2.3	2314	7 US-11-013-759-11	Sequence 11, App1
23	195.5	2.2	1579	7 US-11-052-554A-9	Sequence 9, App1
24	194.5	2.2	5024	6 US-10-793-626-2964	Sequence 2964, App
25	193	2.2	2902	7 US-11-052-554A-91	Sequence 91, App1

26	184.5	2.1	1461	7 US-11-052-554A-283	Sequence 283, App
27	182.5	2.1	3194	7 US-11-052-554A-90	Sequence 90, App1
28	180	2.0	2399	7 US-11-052-554A-92	Sequence 92, App1
29	179.5	2.0	1992	7 US-11-013-759-3	Sequence 3, App1
30	179.5	2.0	1992	7 US-11-013-759-13	Sequence 13, App1
31	179.5	2.0	2047	7 US-11-013-759-4	Sequence 4, App1
32	179.5	2.0	2047	7 US-11-013-759-7	Sequence 7, App1
33	172.5	2.0	5291	7 US-11-052-554A-281	Sequence 281, App
34	172.5	1.9	1279	6 US-10-793-626-3188	Sequence 3188, App
35	172	1.9	1571	7 US-11-052-554A-2	Sequence 2, App1
36	171.5	1.9	1345	7 US-11-052-554A-282	Sequence 282, App
37	170.5	1.9	730	7 US-11-113-424-50	Sequence 50, App1
38	169.5	1.9	2314	7 US-11-097-728-2	Sequence 2, App1
39	168.5	1.9	2340	7 US-11-052-554A-171	Sequence 171, App
40	168	1.9	915	6 US-10-647-956A-6	Sequence 6, App1
41	165	1.9	1095	6 US-10-793-626-3154	Sequence 3154, App
42	165	1.9	2353	7 US-11-097-728-6	Sequence 6, App1
43	163.5	1.8	1588	7 US-11-052-554A-280	Sequence 280, App
44	161.5	1.8	948	6 US-10-485-517-131	Sequence 131, App
45	159	1.8	1126	6 US-10-485-517-248	Sequence 248, App

## ALIGNMENTS

RESULT 1  
US-11-096-051-10  
; Sequence 10, Application US/11096051  
; Publication No. US20050244868A1  
; GENERAL INFORMATION:  
; APPLICANT: Kekuda, Rameah  
; APPLICANT: MacLachlan, Timothy K  
; APPLICANT: Rastelli, Luca  
; APPLICANT: Verne, Corine  
; APPLICANT: Etenberg, Seth  
; TITLE OF INVENTION: Ten-M3 Polypeptides and Polynucleotides and their Methods of Use  
; FILE REFERENCE: Attorney Docket No. Cura 967  
; CURRENT APPLICATION NUMBER: US/11/096,051  
; CURRENT FILING DATE: 2005-03-30  
; PRIOR APPLICATION NUMBER: 10/038,854  
; PRIOR FILING DATE: 2001-12-31  
; PRIOR APPLICATION NUMBER: 10/455,772  
; PRIOR FILING DATE: 2003-06-04  
; PRIOR APPLICATION NUMBER: 60/557,978  
; PRIOR FILING DATE: 2004-03-30  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: CuroSeqList version 0.1  
; SEQ ID NO 10  
; LENGTH: 2721  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-11-096-051-10

Query Match 3.2%; Score 285; DB 7; Length 2721;  
Best Local Similarity 19.5%; Pred. No. 4.4e-10;  
Matches 367; Conservative 253; Mismatches 637; Indels 626; Gaps 97;

18 NEFTQANNFTSAVSGVDPRITGLYNIQITLGHIVGNMIGPTLPPLTSLSPINKTDIGF 77  
1151 NGRTSQPPVWSIMG--NGRRRSICSPSCGQADGNKLLAPVA-----LAC 1195  
78 GI-----GFNGSLVYDRKN--SLTSLS-----TGENYVIE 107  
1196 GIDGSLYVGDYFNVRIPSGNVTSLBLKKNCFRHSNPARYLATDPTDLY--VS 1253  
108 TDTTVKLQCKQDNLRFKDKLKENCRRIIHKSGDIEVLGFNNA--FD-----LKV 157  
1254 DMTTRITVYRK--SLTGADLTGNA-----EVAGTGGQCLPFBARCGDGKAV 1301  
158 PKGLNPAHAIYIDNFEATOPRLNRIYDDLDGHDIPLNLLEYQILKITLTLPQKKE 217  
1302 EATLMSPKGMV-----DKNGLYFVDG--TWIRKVDONGIISTLL-----GSND 1344

```

Qy 218 GYRTLEAFNRQINSIHNFSGENEPYTWSPGYTPIGKNGILQOMITSMAPGAKETVN 277
Db 1345 -----LTSARPLTCD---TSMHISOVRLEMPDL-ALNPMDSIY 1380
Qy 278 YSNNQGHHPQSANLEVLPRVYTLMKQVPG---AGQPAIOAESYTSNHYVGGSGNGIV 333
Db 1381 VLDNNVVLQITENQVRIAGRPMBQVPGVEYEVGKHAQVT---TLESAPALAVSYSGV- 1437
Qy 334 NMLKDNLYGLMTEVNGSTESRRYKDKEGHDQYIRIERTYNNHLLNSECKQON----- 387
Db 1438 -----LY---ITETDEKINRIRQVTTDGEISLVA-----GISECCCKNDANCDC 1480
Qy 388 -----GYQOTE-----TAYVAILGH-----NF-----D 406
Db 1481 YQSGDGAOKAKSLAPSSILASPGDTLYIADLGNIRIRAVSKNPKPLNSNFYVVASPTD 1540
Qy 407 SQPSQFOLPKTKETWBSADNSY-----RSBITETTEDSGNPLTKYIKD---KTKQK 456
Db 1541 QELYIFDINGTHQYVSLVGTGDLNFSYNDNDITAVT---DSNGNTL-RIRRDENRMPVR 1598
Qy 457 IISPSH-WEYYPAGEVUNCPEPYGFTFVKKIIQTPYDSEFKODPEKFIQYRSLSI 514
Db 1599 VVSPDNQVITMLTIGTGCLK-----GMTAQGLEVLFTYH----- 1633
Qy 515 GSQGHVTLKIEBRHYSATQILNSTLFQYNTDKSELGRLKQTECT-----KGNKGTYS 568
Db 1534 GNSGLATKSDPTQW-----TTFPOYDSR---GRITVTFPTQVATMLHGMMDALIT 1682
Qy 569 VVHFYTYTKODDYLQOHSITTHDNF-TIHSQVRSRYTGLFSBDTDTKQIVTQMSYDKL 627
Db 1683 VDIESSSREEDVSI-TSNLSSIDSFYTMVQDOLRNSY-----QIGYDGS 1725
Qy 628 GRLLRTINSTPYANTLT---YDYEINMLQDDNRPPV---ITTDVNGQLRMEFPQAG 682
Db 1726 LRIT-----TASGLDSHYQTEPHVLAGTANPYAKRNMPLGEGONL----- 1768
Qy 683 RHVESQCLSDGDGDKFYTIHTQOYDEQGRHHTSTYSVDLTNGROQTPDKVHLSMSKSYD 742
Db 1769 --VEMRPFKEQAQOK-VAVFGRKLRVNGRNLISVDPORTTBKIKYDHRKFL-LRLAYD 1824
Qy 743 NMGOIANTHWSYGVSEKITTVDPIITLATKQLOSNNSNVQSEKVTYTPSQOPIQITLFD 802
Db 1825 TSGH-PTLML--PSSKLMANVVTYSSYGOIAS---IORG-----TTSER---VDYD 1866
Qy 803 EAGHLQSHLTLPBG--WDRVRKENDAI-----GQCTIYQDNNVRVQITLTP----- 849
Db 1867 GQGRIVS--RVFADGKTWSTYTLKESMWLLHSQKQYIFETDMDRLSAITMPSYAKRTM 1924
Qy 850 -----GTIVRKAPFSTDTLITDIRVNGISLGOQTFDGLSRLTQSGDGRWVAUTYSA 903
Db 1925 QTIHSIGYRNINVPESNASIITDYNREGILL-QTAFIAGTSR-----RV----- 1968
Qy 904 GNDQCPSTVITPDGOFIHYQOPELDDAVLOVASHEITQOQSYNRYVTGALKAV---AEG 960
Db 1969 -----LFKRRQTRLS---ELIYDSTVFSFYDETAG-VIKTVLQSDG 2008
Qy 961 QSLFPIYPSGRLKMEINNDKMSGYLMTLGLNGYTDLTCTIKISRDTHGRTOYK 1020
Db 2009 FICTIRYQOIGPL-----IDKQIFRPSDEGMVAVRD-----YIDNSFRVTSQOG 2054
Qy 1021 SSIKTTL-----NYDDLNRHIGS-----QVTLATGHMLTTTVEFD-GLNREIGR 1064
Db 2055 VINERPLPDLAQFQDISGKVGQFGVYIYDINOIISTAVMTTKFHDHGRKLEIY 2114
Qy 1065 KLCDSGGHTLDIQOSWMLTKQOLANRIYKLVGLQRTQOYSDSRRLNAIOYKCDGACETPD 1124
Db 2115 EIFRSLMWITTIQYV--NMGRVTKREIKI-GPFAVTTKAY-----EYVDG-QLOQT- 2162
Qy 1125 KYGHSIVVQNFYVDIYGNITACHTTPFADQTEDEHAFKFRANPDPQOLTEVHHTHDMDN 1184
Db 2163 VYLNKIKIMRYNYVDNGNL-----HLNPNNSARLTP----- 2194
Qy 1185 IRLKDYKAGRVINITD-----NHGNTENFTYDITGLRLQ-----NGQGSVGYD 1227

```

```

Db 2195 --LRVLDPRITRLGVOVRLDEDFLRORGTEIFEYSSKGLTRVYSKSGVTVIYRYD 2252
Qy 1228 PLNRVLSQRT-----DTLDECELY--RETMLVNEVANGE 1259
Db 2253 GLRRVRSKSLSGOHLQFPYADLTTPRTITRVYNNHSSSETSLYDLOGLPAMEISSGD 2312
Qy 1260 MIRLRTGETIILAQOPASKVLVTGTDQOOSVILTSQKONLSQBAYSAYGR-KHSTANDA 1317
Db 2313 -----EFLIASDN-----TGT---PLAVFSSNGMLKQIQTAYAGBIYFDSNIDFQ 2355
Qy 1318 SILGNGERADPYSGVTHLNGYRSYDPTLMRHTPD-----SLSPFG----- 1360
Db 2356 LVIGFPGGLVDPLTKLHFGE--RDYDILAGRWTTPDIEIWKRIKQDPAPFNLMPERNNN 2413
Qy 1361 -----AGGINPYCYLGDPPIRSND---PSGHL-SWQAWTGIGMI 1396
Db 2414 PASKHVDKDYITDVNSWMLVTFPFLHNLAPGFVFPKFDLTBSYELVKSQOHHDI-PPI 2472
Qy 1397 AGLLRTIATGMAIAAAGIAAIAIAS-----TSTTALAFLALSVTSD---I 1439
Db 2473 FGVQGVARQAKAFLSIGKMAEVOVRRRAGAGQSWLMFATVXSLIGKVMIAVSQGRVQ 2532
Qy 1440 TSIVSGALEDPASPRASSILG-----WVSMGMAAGLAESAIIKGTX 1480
Db 2533 TNYVLIANMEDCI-KVAAVLNNAFYLENLHPTIEBKQDTHYFIKTTTPESDLAGTLRLTSGRK 2591
Qy 1481 LATHILAFADGNNALIKSTSESSRIKMGVTR-----SLDREIVR 1520
Db 2592 AL-----EGINIVTYSQSTIVYNGRTKRPADYEMQFALALHVRIGMTLDEKAR 2641
Qy 1521 NERGOVTKDSRGYTDNFMKGEOAILVHGDKQDFLYHTEGNKN--GK-----GPTT 1571
Db 2642 ILRGARQALABAWA-----REQQVRDGEGBARILM-TGEBKQQLSAGKVQGYDGYV 2694
Qy 1572 RHTPE--QLYDYILKONNITDLTQ 1592
Db 2695 LSVEQYPELAD--SANNIOFLRQ 2715

RESULT 2
US-11-096-051-8
; Sequence 8, Application US/11096051
; Publication No. US20050244868A1
; GENERAL INFORMATION:
; APPLICANT: Kekuda, Rameesh
; APPLICANT: MacLachlan, Timothy K
; APPLICANT: Rastelli, Luca
; APPLICANT: Vernet, Corinne
; APPLICANT: Ettenberg, Seth
; TITLE OF INVENTION: Ten-M3 Polypeptides and Polynucleotides and their Methods of Use
; FILE REFERENCE: Attorney Docket No. Cura 967
; CURRENT APPLICATION NUMBER: US/11/096,051
; CURRENT FILING DATE: 2005-03-30
; PRIOR APPLICATION NUMBER: 10/038,854
; PRIOR FILING DATE: 2001-12-31
; PRIOR APPLICATION NUMBER: 10/455,772
; PRIOR FILING DATE: 2003-06-04
; PRIOR APPLICATION NUMBER: 60/557,978
; PRIOR FILING DATE: 2004-03-30
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Curaseqlist version 0.1
; SEQ ID NO 8
; LENGTH: 2725
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-096-051-8

Query Match 3.2%; Score 285; DB 7; Length 2725;
Best Local Similarity 19.5%; Pred. No. 4.4e-10;
Matches 367; Conservative 253; Mismatches 637; Indels 626; Gaps 97;
18 NEFTQANNFTSAVSGVDPRTGILYNIQITLGIHVGNGLQPTLPLTSLSPLNKTDIGF 77

```



Db 1155 NQFISQPPVSSIMG--NGRRRISCPSCNGADGKLAIPA-----LAC 1199  
 Qy 78 GI-----GPNGLSYDRKN--SLSLSS-----TGNRYVIE 107  
 Db 1200 GIDGSLVYGVJFNVYRRI:FPGSNVTSLVLELNKDFRHSNPAHRYLATDPTDGLY--VS 1257  
 Qy 108 TDKTVKLOQKDLNLFREKDLKENCYRI:HKSGDIEVLGTGPNNA--FD-----LKV 157  
 Db 1258 DMTTRIRYRKK--SLTGAKDLITKNA-----EYVAGTGECLPFDEARCGDGKAV 1305  
 Qy 158 PKKLLNPAHAIYIDWNFATQPLNRIYDDLGDHDIPLNLEYQGLIKITLTFPGQKE 217  
 Db 1306 EATLMSKGMAY-----DKNGLIYFVDS--TWIRKVDNGIISTLL-----GSND 1348  
 Qy 218 GYRTLEAFRLQNLNINPSLGNENPLTWSFGYTPIGKNGILGWITSMAPGSLKETT 277  
 Db 1349 -----LTSARPLTCD--TSMHISQVRLMPTDL-AINEMDSIY 1384  
 Qy 278 YSNNGQHHPPOSANLPLVYVTLMKQVPG--AGPAIOAEXSYSHNYVGGSGSNGIW 333  
 Db 1385 VLDMNVVLQITENQVRIAGRPWHCOVPGVEYVGAHVT--TLBSATLAVSYSGV- 1441  
 Qy 334 NNKLDNLGLMTEYNSTESRRYKDEGHDQIVIRERTYNNYHLLTSECKOON----- 387  
 Db 1442 -----LY--ITETDEKKIRIRQVTTDGEISLVA-----GIFSECDCKNDANCDC 1484  
 Qy 388 -----GYIOTTE-----TAYVAITGH-----NF-----D 406  
 Db 1485 YOSGDGAKDAKLSAPESLASPDGTLIADLGIRIRAVSNKPLLSNMFFEVASPTD 1544  
 Qy 407 SQSPQPLPKTKETWMSADNSY-----RSRITETTESGNPLTKVIKD--KTKQK 456  
 Db 1545 QELYIFINGHOYTSLVGTGDIYNFSYSNDNDITAVT--DSNGNTL-RIRDDNRNRPVR 1602  
 Qy 457 IISPTH--WEYYPAGEVNDNCPPEYGFTRFKVKKIQTPEDSHFQKDPKFIQYRSLI 514  
 Db 1603 VVEPBDNGVILTLTGNGCLK-----GMTAQGLVLFYTH----- 1637  
 Qy 515 GSGSHVTLKIBERHYSATQLNLSTLFOYNTDKSELGRLKQTECT-----KGENKTYIS 568  
 Db 1638 GNSGLLTKSDETEW-----TTFPYDSE--GRLTAVTPFGVAVTMHGMDAIAIT 1686  
 Qy 569 VAKHFTYTKDDTLQOQSHITTHDNF--TIHRSQVRSYTGELFSDPTDITQVOMSYDL 627  
 Db 1687 VDIESSREEDVSI--TSNLSSIDSFYTWODQLRNSY-----QIGYDGS 1729  
 Qy 628 GRLTLRLNAGTPYANTLTL--YDEYLNLODNRPPFV--ITTDVNGQLRNEFDGAG 682  
 Db 1730 LRIT-----YASGLDSHYOTEPHYVLAGTANFYAKKRNMTLPGENGONL----- 1772  
 Qy 683 RHVSQCLKSDGDKFTYTIHQYDEQGRHHTSYSDYLINGRQOTDPDKVHLSMSKSYD 742  
 Db 1773 --VEMRFRKQAOQK--VNVFGRKLRVNGRNLSVDPRRTTKTKIYDHRKFL-LRIAYD 1828  
 Qy 743 NMGQIANTHNSYGVSEKITYDPIITLTKLOQSNNSNVQGEKVTYTTTPSQOPIQITLFD 802  
 Db 1829 TSGH--PTLWL--PSSKLMANVNYSSGTQIAS--IQRG-----TTSK-----VDYD 1870  
 Qy 803 BAGHLOSCHTLTRDG--MDRVKKTDAI-----GQCTIYQDYNRYVQITLTPD----- 849  
 Db 1871 GQGRIVS--RVFADGKWSYTYLEKSNVLLHSQRQIYFEDMDRSLAITMPEVAHNTM 1928  
 Qy 850 -----GTIVNRKXAPSTDTLITDIRVNGISLQOQTGDLISRLTQSGDGRVMAYYTSA 903  
 Db 1929 QTRISIGYRNIVYPPESNASIITDYNEBGLL--QTAFLGTSR-----RV----- 1972  
 Qy 904 GNDQCPSTVITPDCQFIHQYQPLDVAIVOVASNETIQOSYXNPVTGALLKAV--AEG 960  
 Db 1973 -----LFRYRQTLSS--EILDSTRVSTFYDTAG-VLKTNLQSDG 2012  
 Qy 961 QSLTPIYPSGRLKMNENIDMKKMSYLMTLGLNGYVTLGTIQLKISPDTHGHVTOIKD 1020

Db 2013 FICTIRYQIGPL-----IDRQIFRFSEDEGMVNAFD-----YSYDNSFRVTSMOG 2058  
 Qy 1021 SSIKTL-----NYDDLNRHIGS-----QVTLATGMLTTPTEPD--GLNBEIGR 1064  
 Db 2059 VINETPLPIDLYOPDDISGKVEQFGKGVITYDINOIISAVWYTKHFAHRIKIEIYQ 2118  
 Qy 1065 KLCDSGHTLIDIOQSWLKTQOLANRIVKNGVLQRTQOYSYDSRNNLQYKCDGACBCTD 1124  
 Db 2119 EIRFSLMWYITLIQYD--NMGRVYKREKXI--GPEANTTKAV-----EVDVDS--QIQT- 2166  
 Qy 1125 KYGHSIVTONFYTDIYGNITACHTTPADGTEDHATEFKFANPDPCCQITVHHTHPMDPN 1184  
 Db 2167 VYLNEMKIMRYNDLNGNL-----HLNPSNSARLTP----- 2198  
 Qy 1185 IRLKYDKAGVINITD-----NHGNTENFTYDILGRQ-----NGQSGVGYD 1227  
 Db 2199 --LRYDRLDRITBLGDVQYRLDEBGLRQRTSIFEYSSGGLTLRYVSKSGWTVIARYD 2256  
 Qy 1228 PLNRIVSOKT-----DTLDCELYY--RETMVNEVANGE 1259  
 Db 2257 GLGRVYSSKTSLOQHLQFPADLTYPRTITHYNNHSSSETSLYLDQHLFAMEISSGD 2316  
 Qy 1260 MIBLRTGETTIAQORASKVLLGTDSQOSVILTSDKONLSQERYSAYGK--HKSTANDA 1317  
 Db 2317 -----EFYIASDN-----TGT--PLAVESSNGMLKQIQTAYGEIYFDSNIDFQ 2359  
 Qy 1318 SIIYGNERADPPSGVTHLNGRYSYDPTLMRHTPD-----SLSPFG----- 1360  
 Db 2360 LVIGFHGGLVDPPLTKLHFGE--RDYDILAGRWTTPDIEIKRIGKDPAPFNLMPFRNN 2417  
 Qy 1361 -----AGGINPYSYCLDPIRNSD--PSGHL--SWQAWTGIGWGI 1396  
 Db 2418 PASKIDHVKYITDVANSKVLTPFGHLMALPGFPVPEFDLTRESYELVKSQOMDDI--PPI 2476  
 Qy 1397 AGLLTLTANGMAIAAGGIAAIAAS-----TSTTALAFAGLSVTSD--I 1439  
 Db 2477 FGVOQVVARQAKAFLSLGKNAEYQVSHRRAGAQSMLMFAVXSLIKGVTALVWSQVRQ 2536  
 Qy 1440 TSIVSGLEBPASPKASIIIG-----WVSMGMAAGLAEBSIKGTX 1480  
 Db 2537 TNYLANTANECI--KVAAVLNNAFYLENLHPTIGKQTHYPIKTTTPSDIGTLRLTSGRK 2595  
 Qy 1481 LATHLAFADGEGNALKTSSESRIKGVTR-----SLDREIVR 1520  
 Db 2596 AL-----ENGINVTYSOSTVYVNGKTRPADVENQFGALALHVRKMTLDEKAR 2645  
 Qy 1521 NEBGOVICKHSRGYTDNFMKGEOAILMHGDKGFLYHTEGNRN--GK-----GPT 1571  
 Db 2646 ILBOARQALARAFA-----REQQVRDDEBGARLM--TEGKQRLLSAGKVOGYDGYV 2698  
 Qy 1572 RHTPE--QVYDYLKONNIVDLTQ 1592  
 Db 2699 LSVQYPELAD--SANNIOFLRQ 2719  
 RESULT 3  
 US-11-113-424-51  
 ; Sequence 51, Application US/11113424  
 ; Publication No. US20050260713A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gangoli, et al.  
 ; TITLE OR INVENTION: Polypeptides and Nucleic Acids Encoding Same  
 ; FILE REFERENCE: 21402-225  
 ; CURRENT APPLICATION NUMBER: US/11/113,424  
 ; PRIOR FILING DATE: 2005-04-21  
 ; PRIOR FILING DATE: 2000-12-19  
 ; PRIOR APPLICATION NUMBER: 60/311,590  
 ; PRIOR FILING DATE: 2001-08-10  
 ; PRIOR APPLICATION NUMBER: 60/257,314  
 ; PRIOR FILING DATE: 2000-12-20  
 ; PRIOR APPLICATION NUMBER: 60/311,613  
 ; PRIOR FILING DATE: 2001-08-10

```

; PRIOR APPLICATION NUMBER: 60/315,617
; PRIOR FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: 60/307,506
; PRIOR FILING DATE: 2001-07-24
; PRIOR APPLICATION NUMBER: 60/322,358
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/294,075
; PRIOR FILING DATE: 2001-05-29
; PRIOR APPLICATION NUMBER: 60/288,153
; PRIOR FILING DATE: 2001-05-02
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 51
; LENGTH: 2715
; TYPE: PRT
; ORGANISM: Mus musculus
; US-11-113-424-51

Query Match      3.2%; Score 284.5; DB 7; Length 2715;
Best Local Similarity 19.6%; Pred. No. 4,7e-10;
Matches 302; Conservative 214; Mismatches 524; Indels 497; Gaps 80;

QY      18 NEFTQANNFSAVSGVDPRFTGLNITQITLGHIVGNGNLCPPLPLTLSTSPLNKTDIGF 77
DB      1145 NQFISQOPPVVSSIMG--NGRRRSISCPSCNGADGNLAPVA-----LAC 1189
QY      78 GI-----GFNPGLSVYDRKN--SLTSL-----TGENYKVI 107
DB      1190 GIDGSLVGDVFNRYRRIFFPSGNTSVLELRNKRPHHSNPMHRYLATLDPVTGDIY--VS 1247
QY      108 TDKTVKLOQKLDLRFKEDLKENCYRIHKSGLIEVLTFGNNA--FD-----LKV 157
DB      1248 DNTNRIRYRPR--SLTGAKDLTKNA-----EVAAGTGEQCLPDEARCGDGKAV 1295
QY      158 PKKLNPAGHAIIYDWNFEATQPLRLRIYDDLDEHDPLNLBEYQGLIKITLTFPGQKE 217
DB      1296 EATLMSFGMAI-----DKNGILYFVDG--TWIRKYDQNGIISTLL-----GSND 1338
QY      218 GYRTFLRLNQLNSIHNFSLGNEPLNWSFGYTPIGKGLIGOMITSMTPAPGGLKETVN 277
DB      1339 -----LTSARPLTCD--TSMHISQVBLEPFDL--AINPMDSIY 1374
QY      278 YSNNNGHFPQSANLPVLPLYTLMKQVPG---AGQPAIQAEYSYSHNYVGGGSGNGIC 333
DB      1375 VLDDNNVVLQITENQVRIAAGRPMHCQVPGVEYPRGKAQVQT--TLBSATAIAVSYSGV- 1431
QY      334 NNKLDNLGLMTENVYSGTESRRYKDKGHOIYRIETNNYHLLTSECQON----- 387
DB      1432 -----LY--ITETDEKKIRIRQVLTGTDEISLVA-----GIPSECCCKDANDC 1474
QY      388 -----GYIOTE-----TAYYAIIGH-----NF-----D 406
DB      1475 YQSGDGYAKADKALNAPSLASPQGTLYIADIGNRIRAAVSKKMLNSMFEVYASPTD 1534
QY      407 SQPSQFOLPKTKTETWRASANSY-----RSEITETTFDESGNPLTKVIKD--KKTQK 456
DB      1535 QELYFIDINGHQYVLSLVGTGYLVNFSYNDNDVTAVT--DSNGNTL--RIRDDPRMREVR 1592
QY      457 IISPEYTH--WEYYPAGEVYDNCPPPYGTRPVKKIIOQTPDSEKDKPEKFIQRYSLI 514
DB      1593 VVSPDNQVIMTLTGTNGCLKSMTAGL-----BLVLFTH-- 1627
QY      515 GSQSHVTLKIEBRHSATQNLNSTFYQYNTDKSELGRLLKQECT-----KGENGKTY 568
DB      1628 GNSGLATKSDETG-----TTFPDVDF-----GRLLNTFPFGVNTLHGMDKAIT 1676
QY      569 VVHKETTYTKODDTLOQSHSITTHNF--TIHSQVRSRYTGRLLSPTDKDITVQMSYKL 627
DB      1677 VDIESSREEDVSI--TSNLSSIDSFYTMVQDLNSY-----QGYDGS 1719
QY      628 GRLLTRLTNSGTPYANLTL--YDYELANLQDNRPPV---ITTDVANGQLRAEFDDAG 682
DB      1720 LRIF-----YASGLDSHYQTEPHVLAGTANPVAKRMTLTPGNGONL----- 1762
```

```

QY      683 RHVSQCLKQSDGDKGKFTYITHTQYDEGRHHTSTYSYLTLNGRQOTDPRKVLHSMKSYD 742
DB      1763 --VEMRFRKEQAGK--VNVFGRKLAVNGRLLLSVDFPRTTKTEKIYDHRKFL--LRAYD 1818
QY      743 NWQILANTHWSYGVSEKIVDPIITLTKAQLOSNNSNVQKGAVTTTYSQOPIQITLFD 802
DB      1819 TSGH--PTLWL--PSSKLMAVNTYSSTQIAS-----IQRG-----TTSEK-----VDYD 1860
QY      803 EAGHLQSCHTLTBDG--WDRVRKETDAI-----GQCTIYQDYNRYIOQLTLPD----- 849
DB      1861 SQGRIVS--RVPADGKTSYTYIEKSNVLLHSQROTYIEYDMMDRLSALITMPSVANHTM 1918
QY      850 -----GTIVNRKXAPSTDTLITDIRVNGISLGOQTFDGLSRLOSQDGRVWATYISA 903
DB      1919 QTRISGYVNRININPPESNNSIITDVNEGLL--QTFALGTSR-----RV----- 1962
QY      904 GNDQCESTVTPDQGFHYQYQPELDVAIQVANSNETQOFSNVPVTKALLAV--AEG 960
DB      1963 -----LEKYSRQTRLS-----EIIYDSTRVSFTYDETAG--VLKTVNLQSDG 2002
QY      961 QSLTPIYPSGRUKMENINDMKMSYLMTLRGLENGVTDLTGTLQKISRDPHGRVTOIKD 1020
DB      2003 PCTIRYRQIGPL-----IDQIRFSEDEGMNARFD-----YSYNSFRVTSMQG 2048
QY      1021 SSIKTYL-----NYDDLNRHIGS-----QVTLATGMLTLYVEFD--GLAREIGR 1064
DB      2049 VINETPLPIDLYQFEDISGKVEQKRGVIYDINOIISTAVMTYTGHPAHGRIKEIYQ 2108
QY      1065 KLCDSGHTLTDIQOSWLTKOQLANRYKLVGLVLORTQYSDSNNRNLQYKCDGAEPTD 1124
DB      2109 EIFRSLMYWTITQYD--NMGRVYKREIKI--GPANITIKYAV-----EYDVG--QLOT- 2156
QY      1125 KYGHSIVTQNFYDIQNTIRACHTTFPADGTEDHATFKAPATPDCOLTEVYHHTPMPDN 1184
DB      2157 VYLAKEKIMRKYNDLGNL-----HLMPSSARLTP----- 2188
QY      1185 IRLTYKRGAVINITD-----NHGNTENFYDTLGRLO-----NGQGSVYGYD 1227
DB      2189 --LRYDLRDRITRGLDQVYRLDEDEGFLRQGTETFEYSKGLLRVYSKSGMGTVIYRYD 2246
QY      1228 PLNRVYQKT-----DTLDCELY--RETMUNVEVRNGE 1259
DB      1260 MIRLRTGETLIAQGRASKVLTLGTDSQGSVILTSDQNLQSOEYASNGK--HKSTANDA 1317
DB      2307 -----EFTIASDN-----TGT---PLAVSSNGLMKQOTYATGEIYPPSNVDFO 2349
QY      1318 SILGNGERADPVSGVTHLNGVRSYDPTLMRFTHPD 1354
DB      2350 LVIGFHGGLVDPLTKLHFGE--RDYDILAGRWTTPD 2384

RESULT 4
US-11-096-051-4
; Sequence 4, Application US/11096051
; Publication No. US20050244868A1
; GENERAL INFORMATION:
; APPLICANT: Kekuda, Ramesh
; APPLICANT: MacLachlan, Timothy K
; APPLICANT: Raetelli, Luca
; APPLICANT: Verneil, Corine
; TITLE OF INVENTION: Ten-M3 Polypeptides and Polynucleotides and their Methods of Use
; FILE REFERENCE: Attorney Docket No. Cura 967
; CURRENT FILING DATE: 2005-03-30
; PRIOR APPLICATION NUMBER: 10/038,854
; PRIOR FILING DATE: 2001-12-31
; PRIOR APPLICATION NUMBER: 10/455,772
; PRIOR FILING DATE: 2003-06-04
; PRIOR APPLICATION NUMBER: 60/557,978
```

PRIOR FILING DATE: 2004-03-30  
 NUMBER OF SEQ ID NOS: 38  
 SOFTWARE: Curaseqdist version 0.1  
 SEQ ID NO 4

LENGTH: 2376  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-11-096-051-4

Query Match 3.2%; Score 283; DB 7; Length 2376;  
 Best Local Similarity 19.4%; Pred. No. 4.8e-10;  
 Matches 365; Conservative 255; Mismatches 637; Indels 626; Gaps 97;

18 NEPTAANFTSAVSGVDPTGLYNIQTILGHIVGNGPTLPILTSYPLNKTDIGF 77  
 806 NQPTSQPPVVSIMG--NGRRRSISCPSCGADGKNKLAPVA-----LAC 850  
 78 GI-----GFNGLSYVDRKN--SLSLSS-----TGENYKVIIE 107  
 851 GIDGSLVGGPNTVRRIFPSGNVTSVLELNKDFRHSNPAHRYLATDPTGDLV--VS 908  
 108 TDKTVKLOQKLDNLAREKDKENCYRIIHKSGDIEVLTFGRNNA--FD-----LKV 157  
 909 DTYRRRIYRPK--SLTGAKDLTKNA-----EYVAGTGEQCLFDEARCGGKAV 956  
 158 PKLLNAGHAIIYIDMNFATQPRLNRIYDDLGDHDIPLNLEYGILKITLLTFPGKE 217  
 957 EATLMSKGNV-----DKNGLIYFVDS--TWIRKVDNGIISTLL--GSND 999  
 218 GYRTELFLNQLNSIHNFSIGNENPLTWSFGYTPIGKNGILGOMITSMAPGGLKETVN 277  
 1000 -----LTSARPLTCD--TSMHISQVLEWPTDL--AINPMDSIY 1035  
 278 YSNNNGHFPPOSANLFLPYVTLMKQVPG--AGCPAIQAEYSYTSNHYVGGSGNGIW 333  
 1036 VLDNNVVLQITENQVRIAGRPNGCQVPGVEYVGGHVAQT--TLESATAIIVSYSGV- 1092  
 334 NNKLDNLXGMLTEYNSTESRRYKDKEGHDQYRIERTNNHLLNSECQON----- 387  
 1093 --LY--ITETDEKKINRIKOVTTDGEISLVA-----GIPSECDCKNDANCDC 1135  
 388 --GYIQTTE-----TAYVAIGH-----NF-----D 406  
 1136 YQSGDGYAKDAKSLAPSLSAPDGTLYIADLGNIRIRAVSKNPKLNSNFYVVASPTD 1195  
 407 SQPSQFOLPKTETWRSADNSY-----RSIETTPDESQNPULTKVIKD--KTKQK 456  
 1196 QELTFPINGHQYTVSLVTGDLYNFSYSNDNDITAVT--DSNGNTL--RIIRDENRMPVR 1253  
 457 IISRSTH--WEYYPAGEVUNCPEPYGFTRFVYKIIITPYDSFKODPEKFIORYSLI 514  
 1254 VVSPDNOVIMLTIGTNGCLSMTAQGL-----ELVLFYTH-- 1288  
 515 GSQSHVTLKIEERYASATQLNSTLFOYNTDKSELGRLLKQTECT-----KGENGTYS 568  
 1289 GNSSLATKSDETCM-----TTFPYDSR-----GULTNVTFTYGVVTHLHGMDAIIT 1337  
 569 VHKFTYTKODDTLQOQSHSTTHDNF--TIHSQVRSRYTGLFSDTDTKDIIVTQMSYDKL 627  
 1338 VDIESSREEDVSI--TSNLSISDSFYTMVQDQLRNSY-----QIGDGS 1380  
 628 GRLLTRLNLSGTTPYANLTL--YDEELNNLQDDNRPVY--ITTDVNGNQLNRPFGAG 682  
 1381 LRIT-----YASGLDSHYQTEPHVLGATANPYAKRNMTLPGENGQNL----- 1423  
 683 RHVSQCLKSDGDKGFTYIHTQOYDEQGRHHTSTYSYDLTNGRQOTPDKVHLSMSKSYD 742  
 1424 --VEMRRFRKGAQOK--VNVPRKRLAVNGRNLSVDFPRTTTEKIYVDHDKFL--LRAYD 1479  
 743 NMGQIANTHMSYGVSEKITVDPTLTATKOLQSNNSNVQGEVTTYTPSQOPIQITLFD 802  
 1480 TSGH--PTLML--PSSKLMVAVNTYSSTGQIAS-----IQSG-----TTSKK-----VDYD 1521

803 EAGHLSQCHTLTRDG--MDRVKRTDAI-----GQCTIYQDVNRYNIQITLFD----- 849  
 1522 GQGRIVS--RVFADGKMTASTYLEKSMVLLHSQRQRIEYEDWMDLSAITSVARHTM 1579  
 850 -----GTVNRKXAPPSDTTLTDIRVNGISLGQOTPDGLSFLTOSQDGRVMAITYSA 903  
 1580 QTRISIGYNNIINYPESNASIITDYNEBGLLL--QTAFLGTSR-----RV----- 1623  
 904 GNDQCSVTTPGQGIHQOYQPELDVAVLQVANSNIITQGSNYPVTGALLKXAV-----AEG 960  
 1624 -----LFTKRQTRLS-----EILYDSRVSFYDETAG--VLKTVNLQSDG 1663  
 961 OSITPIYSGRLKMNINMDKMSYLTLLRGLNGEYDTLGTITOKISRDPHGRVYQIKD 1020  
 1664 FICTIRYQGLP-----IDQIIFRPSBDGVNNAFPD-----YSYDNSFRVTSMG 1709  
 1021 SSIKITL-----NYDDLNRHIGS-----QVTLANGHMLTTVEPD--GLNREIGR 1064  
 1710 VINETPLPIDLYQPDISGKVEQFGKFGVLYYDINOIISTAVWYTKHFPDAGRIRKEIQY 1769  
 1065 KLCDSGHTLDDIQSWLKTQQLANRYKLVKNGVLRTEQYSYDSRNLNLYQKCDGACPTD 1124  
 1770 EIPRSLMYWITIQYD--NNGRVTKRRIKI--GPPANTTKXAY-----EYVDG--QLQT-- 1817  
 1125 KYGHSIVTQNFYDIGNITACHTTADGTEDEHATKFPANPTDPCOLTEVHTHPMPDN 1184  
 1818 VYLNKIMRKYNDIANGNL-----HLNPSASALT-- 1849  
 1185 IRLKYDPAKRVINITD-----NHGNTENFTYDTLGRLO-----NGQGSYVGYD 1227  
 1850 --LRYDLRDRITRLGQVRYRLDEDFLRQGTETFEYSKGLLTRVYSKSGSVTIYRYD 1907  
 1228 PLNRVYSQK-----DTLDCELY--REIMLVNVRNGE 1259  
 1908 GLRRRVSKTSIGQLHOFYADLTPTPTRIHYVNHSSSEITSLYLDLQGLFAMEISSG 1967  
 1260 MRLRTGETIIAQORASKVLGTGDSQGSVILTSQKONLSQEAAYAGK--HKSTANDA 1317  
 1968 -----EYIASDN-----TGT--PLAVSSNGMLKQIQTAYGSEIYDSDNDFQ 2010  
 1318 SILGNGERADPVSGVTHLNGVRSYDPTLMRPHTD-----SLSPFG----- 1360  
 2011 LVIGFNGGLYDPLTKLIHFGR--RDYDILAGRWTTDIEIKRIGKDPAPFNLVMEFANN 2068  
 1361 -----AGINPYSYCLGPINRSD--PSGHL--SMQANTGICMG 1396  
 2069 PASKIHVDQYITDVNSWLVTFGFHLNALPGFPVPEKFDLTESYELVYSQOWDDI--PPI 2127  
 1397 AGLLTIATGMAIAGAIAAIIAS-----TSTTALAFAGLSTSD--I 1439  
 2128 FGVOQOVARAKAFSLIGKMAEYQVSRBRAGAQSWMFATVYSLIGKGYMLAVSQGRVQ 2187  
 1440 TSIISGALBDASPKASSILG-----WVSMGMAAGLAESAIGKGTK 1480  
 2188 TNYLTANANEDCI--KVAALVNANAFYLENLHTIEKGOTHYIKTTPESDGLTRLTGRK 2246  
 1481 LATHLGAFAEDGENALIKTSBSSRIKMGYTR-----SLDEIYR 1520  
 2247 AL-----ENGINVTYOSITVYNGRTTRPADVENQFGALALHVRGMLTDEKAR 2296  
 1521 NBSGOVYKDSRGYTNFMQGEQALIVHGDKXGFLYHTEGNGKN--GK-----GPLYT 1571  
 2297 ILEQAKORALAKMA-----REQQRVDSDEBARKLM--TDEGRQLLSAGKVGQGYDGYV 2349  
 1572 RHPPE--QLVDYLDKNNIYDLTQ 1592  
 2350 LSVQYQPELAD--SANNTIGLRQ 2370

RESULT 5  
 US-11-096-051-2  
 ; Sequence 2, Application US/11096051  
 ; Publication No. US20050244868A1

GENERAL INFORMATION:  
APPLICANT: Keshda, Ramesh  
APPLICANT: MacLachlan, Timothy K  
APPLICANT: MacLachlan, Luca  
APPLICANT: Eitenberg, Seth  
APPLICANT: Eitenberg, Seth  
TITLE OF INVENTION: Ten-43 Polypeptides and Polynucleotides and their Methods of Use  
FILE REFERENCE: Attorney Docket No. Cura 967  
CURRENT APPLICATION NUMBER: US/11/096,051  
CURRENT FILING DATE: 2005-03-30  
PRIOR APPLICATION NUMBER: 10/038,854  
PRIOR FILING DATE: 2001-12-31  
PRIOR APPLICATION NUMBER: 10/455,772  
PRIOR FILING DATE: 2003-06-04  
PRIOR APPLICATION NUMBER: 60/557,978  
PRIOR FILING DATE: 2004-03-30  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: Curoseqblast version 0.1  
SEQ ID NO 2  
LENGTH: 2715  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-11-096-051-2

Query Match 3.2%; Score 283; DB 7; Length 2715;  
Best Local Similarity 19.4%; Pred. No. 5.9e-10;  
Matches 365; Conservative 255; Mismatches 637; Indels 626; Gaps 97;

18 NEFTQANNFTSAVSGVDPRTGLYNIQITLGHIVGNCMLPPLTSLYSPLKNTDIF 77  
145 NQFISQCPVVSISMG-NGRRRSISCPSCGQADGNKILAPVA-----LAC 1189  
78 GI-----GNFGLSVYDRKN--SLSL-----TGENYKYE 107  
1190 GIDGSLVYGDRYVARIPESSGNTSVLELRNKDFHSSNPAHRYLATDPVTDLY--VS 1247  
108 TDKTYKLOOKLDLNFPEKDLKENCYRIHKSGLIEVLTFGNNA--FD-----LKV 157  
1248 DTRNRIIRPK--SLTGAKDLTKNA-----EVAGTGEQCLPDEARCGDGGAV 1295  
158 PKGLINPAHAIYIDWNEFATQPRINRIYDLDGHDIFLNLDEYQGLIKITLTPPGKE 217  
1296 EATLMSPKGMV-----DKNGLIYFVDG--TWIRKVDQNGIISTLL-----GSND 1338  
218 GYRTLERLANQNLNINPISGENEPLTMSGYTPIGKNGILGOWITSMAPGKJKEVFN 277  
1339 -----LTSARPLTCD--TSMHISQVRLMPDCL-AINPDNSIY 1374  
278 YSNNOGHNFPOSANLPVLPVTLMKQVPG--AGOPAIQAEYSYSHNYVGGSGNGIM 333  
1375 VLDDNNVVLQITENRQVRIAAGRPNCQVPGVEYVPGKHAQOT--TLBSATAMAVSYSGV- 1431  
334 NKKLDNLVGLMTEVYSGTESRRYDKGHDQIVRIERTYNNYHLLTSECQON----- 387  
1432 -----LY--ITETDEKKINRIQVTTDGEISLVA-----GIPSECCCKDANDC 1474  
388 -----GYQTE-----TAYVAIIIGH-----NF-----D 406  
1475 YOSGDGVAKDAGLAPSSSLASPDGTLIADIGNRIRAVSKKPLNSMNFYEASPTD 1534  
407 SOPSOFLPKTKETETWRSDNSY-----RSEITETTFDESIGNPLTKVIKD--KTKQK 456  
1535 QELYFDINGTHQYVSLVTDGYLNPFSYNDNDITAVT--DSNGTL--RIRDRPMBVR 1592  
457 IISPETH--WEYYPAGEVNDCRPEPYGFTFRVKKI IOTPYDSEKDPKFIQYRYSLI 514  
1593 VVSPNQVILWLTITNGCLKSWTAQGL-----ELVLETTYH-- 1627  
515 GSGSHVTLKIEBRHSAQOLNSTLFOYNTDKSEHGRLLKQTECT-----KGENGKTVS 568  
1628 GNSGLLAKTSDETEG-----TTFPDYDS--GRLLNVTFPGVVNTNLHGMDKAIT 1676  
569 VNHKFTYTKQDDTLQOSSHITTHDNF--TIHRSQVRSRYTGRLLFSDTDKDIYTGMSYDKL 627

1677 VDIESSEREEDVSI--TSNLSSIDSFYTWVQDDLRNSY-----QIGVDGS 1719  
628 GRLLTRTLNSGTYRANTLT--YDELANLQDDNRPPV-----ITTDVNGQLRNRPFGAG 682  
1720 LR11-----YASGLDHSYQTEPPVLAGTANPFAKKNMMLPGSEGNL----- 1762  
683 RHYSQCLKSDGQKFFTHITQOYDEQGRHNTSTYSYDLTNGRQQTDPDKVHLSMSKYD 742  
1763 --VEMRRKEQAOQK--VNVGRKLAVNGRLLSVDPRTTKTEKIYDHRKFL--LR1AYD 1818  
743 NMQIANTHWSYGVSEKITYDPIITLATAKQLQSSNNVQGEVTTYTPSQOPIQITLFD 802  
1819 TSGH--FTLWL--PSSRLMAVNVYSSTGIAS-----IQRG-----TSEK-----VDYD 1860  
803 EAGHLOSCHTLTDG--WDRVKETDAI-----GQCITQYDNNRYIQTLPD----- 849  
1861 GQGRIVS--RVFADGKTSYTYLEKSNVLLHSORQYIFEDWMDRLSALTMPSVARHTM 1918  
850 -----GTIVNRKVAPESTDITLDIRNGISLGOQTFDGLSRLTQSDGGRVMAVYISA 903  
1919 QTRISGYRYRNINPSSNASIITDYNBGLL--QTAFLETSR-----RV----- 1962  
904 GNDQCPSTVITPQGFTHQYQPELDAVLQVANSNEITQGFSPYNPTGALLKAV--AEG 960  
1963 -----LFTYRRQTRLS---EILYDSTRVSFTYDETAG-VLKTVNLQSDG 2002  
961 QSLPTIYPSGRKLMENINDMKMSYMLTARGLENGTDLTGTCQKSRDTHGAVTQIKD 1020  
2003 FICTIRRAQIGPL-----IDRQIFRFSBDGMVNAFDP-----YSNDSFRTVMQ 2048  
1021 SSIKTL-----NYDDLNRHIGS-----QVTDATGMLTLTYEFD--GINREIGR 1064  
2049 VINETPLPIDLXPDDLSGKVEQGRGVYYIDNOIISRAVMTYTHGFPAHGRIKIYQ 2108  
1065 KLCDSGHTLIDIQOSWLKTOQLANRIKLVNGVLQRTQOYSYDSNRNLNQYKCDGAEPTD 1124  
2109 EIFRSLMWITIQVD--NMGRVTKREIKI--GPFANTTKYAV-----EYVDG--QLOT-- 2156  
1125 KYGHSIVTONFYDNYGNITRACHTTFADGTEADATFPFAPDPCQLTENVHTHPDMPDN 1184  
2157 VYLNKIKIMWRKYNVDLNNL-----HLNPSNARLTP----- 2188  
1185 IRLKYDKRAGRYINTD-----NHGNTENFTYDLAGLQO-----NGQGSVYGYD 1227  
2189 --LRYDLRDKITRIGDVQYQLDEBDFLRQGTETEFESSKGLLRVYSKSGMWIYIRYD 2246  
1228 PLNLVSOQT-----DTLDCELY--RETMVLNVEVRNGE 1259  
2247 GLGRVSSKTSLQHLQFYADLTYPTRITHVYHSSSEITSLYYDLQGHLPAMEISBGD 2306  
1260 MIRLKTGERTIANQARSKYLCTGDSQASVILTSDQONLSQEAYSVAGK--HKSTANDA 1317  
2307 -----EFTYADN-----TGT--PLAVFSSNGMLKQIOYTAIGELEYFSPSNDIFQ 2349  
1318 SILGNGERADPVSGVTHLNGVYSYDPTLMKRFPTD-----SLSPFG----- 1360  
2350 LVIGFHGLVDPRLKLIHFG--RDYDIAGRWTPPIEIKRKIGKOPAPFNLMPFRNN 2467  
1361 -----AGGINPYSYICGDDPINSD--PBGHL--SWQATGIGMGI 1396  
2408 PASKIHVKDYITDVNSMLVTFGFHLNNAIPGFVPKFDLPEPSYELVKSQOMDPI--PPI 2466  
1397 AGLLTITATGMAIAAAGIAAALAS-----TSTALAFGALSVSD--I 1439  
2467 FGVOQOVARQAKALSLGKMAEVOVSRRAAGASQSWIMFATVSKLIGGVMLAVSQGRVQ 2526  
1440 TSVISGALSDAPSPASSILG-----VWSMGMAAGLAESAIKGCTX 1480  
2527 TTVANTIANEDCI--KYAAVLNNAFYLENLHPTIBGKQTHYFIKTTPPSDGLTALTSRK 2585  
1481 LATHIGAFAEDGENALLKSTSESGRIKGVTR-----SIDREIVR 1520

Db 2586 AL-----ENGINTVQSSTVVNGFRPRADVYMGFGALAHVYGMATLDEEKAR 2635

Qy 1521 NEEGVITKDSRGYTDNFMKGEOALIVHGQDGLFVITBENKIN-----GK-----GPYT 1571

Db 2636 ILLEGARGALLARMA-----REGORVADGSEGRALW-TBEEKRQLLSAGNYGVYDGYTV 2688

Qy 1572 RHTE--QVYDYLDKDNINVDLTQ 1592

Db 2689 LSVGEYFELAD--SANNIQFLKQ 2709

RESULT 6  
IIS-10-A2

```

Sequence 1097, Application US/10821234
Publication No. US20050255114A1
GENERAL INFORMATION:
APPLICANT: Labat, Ivan
APPLICANT: Stache-Crain, Birgit
APPLICANT: Andarmani, Susan
APPLICANT: Tang, Y. Tom
TITLE OF INVENTION: Methode for diagnosis and Treatment of Preeclampsia
FILE REFERENCE: 821A
CURRENT APPLICATION NUMBER: US/10/821,234
CURRENT FILING DATE: 2004-04-07
PRIOR APPLICATION NUMBER: US 60/462,047
PRIOR FILING DATE: 2003-04-07
NUMBER OF SEQ ID NOS: 1704
SOFTWARE: pt SEQ_genes Version 1.0
SEQ ID NO 1097
LENGTH: 1094
TYPE: PRT
ORGANISM: Homo sapiens
US-10-821-234-1097

```

Query Match	3.1%;	Score 273.5;	DB 6;	Length 1094;
Best Local Similarity	20.4%;	Pred. No. 5.7e-10;		
Matches 255;	Conservative 164;	Mismatches 417;	Indels 413;	Gaps 62;

```

Qy 594 FTIHRSGVARSRYTRLFSDDTOKOIVQMSYDKGLLTRLTNSGT-----PYANLTLY 647
Db 3 PTHYGN-----SGLLTKSDSETGWTTFPDYDSERGLTNVTFPGVGVTNLHGMDKAITV 56
Qy 648 DYELNLIQDNNRPPEVITTDVNG-----NQLRNE-----FDGAGRHSQCLKDS-- 692
Db 57 DISSSSREED-----VSITSNLSSIDSEYTWQQLRNSYOIGDGSRLRIYASGLDSHY 111
Qy 693 -----DDDKFYTHHQYQDEGHHSTYSDVLTNGR----- 725
Db 112 QTEBHVLAGTANPLTVAKRNMTLLPEBNGONLWEMPRKEQAAQKXNVNPGKRLVNGRNILS 171
Qy 726 -----QOTDPDKVHLMSMK-----SYDNWGOJANTHWSYGVSEKTYVPIITLATAKOLQSN 776
Db 172 VDPDRTKTEKXIYDHRKFLRLRIAYDPSGH--PLML--PESKLMANVVYTSSTGQLAS-- 226
Qy 777 SNNVQTEKEVYTTYPPSQOPIOTILPDEAGHLQSCHILTRDG--MDRVRKFTDAI-----G 829
Db 227 ---IQRG-----TTSSEK-----VDYDQGRIVS--RFAPADKWTSSYTLLEKSMVLHLSQ 271
Qy 830 QCTIYQYDNNRYVQITLPD-----GTIVRKVAPFSDTILITIRVNGSISG 877
Db 272 ROYIFEBIDMMDRLAITMPSVARHTMQIRISIGYRRIYNPNPESNASTIIDYNEEGILL-- 330
Qy 878 QCFPDGLSKLTQSDGGRWMAVYTTASAGNDQCPSTVITPDGOFIHYQVQPELDNAVLOVAS 937
Db 331 QTAFLGTSR-----RV-----LFKRRRTRLS-----ELLY 356
Qy 938 NEITQCSYNPVNTGALLKAV---AEGOSLPYIYPSGRLKMENTDMKMSYLTWLGLE 994
Db 357 DSTRVSTYDETAG-VLKTWVLQSGDFICTIRYQIGSL-----IDRQIFRSEBGMV 408
Qy 995 NGYTDLTGTLQIKSRBDTHGRVTOIKDSIKTTL---NYDDLNRHGIS-----QV 1040
Db 409 NARDP-----YSDNSFRVTSNQGVLNPNPLPIDLVQPDFIDISGKEQFQPKGVIIYDI 461

```

Qy	1041	TDLATGMLTTFVAFD--GANREIRGLCCSSGGHTIDIQOSMLTKTOOLARIYVLTNGVLQ	1098
Dd	462	NQIISTAVMTYTHKFDHNGRIKKEIQEIPFSLMTWTITQYD--NMGRVTHGREIKI--GPA	518
Qy	1099	RTEQYSYDSRNRNLNQYKCDGACBCEPTDKYGHISVTQNTFYDIYGNITPACTHTFADGTEDHA	1158
Dd	519	NTTKYAY-----EYDVGD-QLOT--VYLNKEIKIMWRNYVDLNGUL-----	554
Qy	1159	TFKPAKNTDCCQLETVHHTHPMDNRILKYDAGRIINITD-----NHGVTEN	1207
Dd	555	--HLMNSNSARLTP-----LRTDLRDRITRGLDVQYRLDEDFLRQRGTEI	599
Qy	1208	FTYDILGRLO-----NGGQSVYGYDPLNLVSOXT-----	1237
Dd	600	FVYSKGLLFRVYKSGSGMTVIVRYDGLGRVASKTSLGHLQEFVADLTYPRITHVYN	659
Qy	1238	--DTLDELTY--RETMVNEVNGEMIRLLRGETIILAOQPAKULLTGTDSQOSVILT	1293
Dd	660	HSSEITSLLYDLOGHLFAMEISSD-----EPIYASDN-----TGT--PLAVFS	702
Qy	1294	SDKONTSQEAAYSAYGK--HKSTANDASILGNGERADPVGVTILGNGYRSYPTLKRFX	1351
Dd	703	SNGMLKQIQYTAIVGELIYFDSNIDFQLVIFHGLVYDPLKRLHPGE--RDYDILAGRWT	760
Qy	1352	TPD-----SLSPFG-----AGINPVSICYGDP	1374
Dd	761	TPDIEIKRIGKDPAPFNLVYFRNNNPASKIHVDKDYITDVNSWLVTYFGHLHNAIDGFP	820
Qy	1375	INRSD---PSGHL-SWQAWTGIGMGIGLTLTATGMLAAAGIAPALAS-----	1422
Dd	821	VPKRDLTEPSELYLVKSQOMDI--PPIFGVQOQVAPAROKAKFLSLCKMALEVQVSRBRAGAQ	879
Qy	1423	-----TSTTALAFGLSVTSD--ITSIVSGALEDSPPASSILG-----	1459
Dd	880	SWLFAFVVKSLIGKGVWLAVSOGRVQTNVIANINEDCI--RYAAVLNNAFYLENLHFTTIEG	938
Qy	1460	-----WTSWMGGAAGLAEASAIKGGTYKATLHGLAFEDGENAALLKSTSESRIKMGVTR--	1512
Dd	939	KOTHYFPIKTTTPESDLGTLRLTSGRKL-----ENGINVTYSQSTTVVNGRRTRRF	988
Qy	1513	-----SLDREIVRNEBQVIKDHSRGYTDNPFMGKGEQALLVHGDKG	1554
Dd	989	ADVEMQEGALALHVRVGMTEDEEKARILFQARQALARAFA-----REGQRYADGBEGA	1042
Qy	1555	FLVHTBEKNEN-----GK-----GPRYHTEB--QVUDYLLKDNNTVDTLQ	1592
Dd	1043	RLW--TBEKRQQLSAGKRVQGYDGYVLVSVQEPBLAD--SANNIQFRQ	1088

## RESULT 7

US-10-453-372-170  
Sequence 170. Application US/10453372  
Publication No. US2006000323A1  
GENERAL INFORMATION:  
APPLICANT: Alsbibook, et al.  
TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHODS  
FILE REFERENCE: 21402-589 A  
CURRENT APPLICATION NUMBER: US/10/453,372  
CURRENT FILING DATE: 2003-06-03  
PRIOR APPLICATION NUMBER: 09/789390  
PRIOR FILING DATE: 2001-02-23  
PRIOR APPLICATION NUMBER: 60/185967  
PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: 09/823187  
PRIOR FILING DATE: 2001-03-29  
PRIOR APPLICATION NUMBER: 60/195792  
PRIOR FILING DATE: 2000-03-10  
PRIOR APPLICATION NUMBER: 09/839446  
PRIOR FILING DATE: 2001-03-19  
PRIOR APPLICATION NUMBER: 60/199476  
PRIOR FILING DATE: 2000-03-25  
PRIOR APPLICATION NUMBER: 09/863776

PRIOR FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: 60/208263  
PRIOR FILING DATE: 2000-05-31  
PRIOR APPLICATION NUMBER: 09/939398  
PRIOR FILING DATE: 2001-08-24  
PRIOR APPLICATION NUMBER: 60/227800  
PRIOR FILING DATE: 2000-08-25  
Remaining prior application data removed - See file wrapper or PAM.  
Number of SEQ ID NOS: 1609  
SOFTWARE: Cureseqlet version 0.1  
SEQ ID NO 170  
LENGTH: 2333  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-453-572-170

Query Match 3.1%; Score 272.5; DB 6; Length 2333;  
Best Local Similarity 19.5%; Pred. No. 2.3e-09;  
Matches 365; Conservative 251; Mismatches 652; Indels 603; Gaps 99;

49 GHVING-NIGPTPLPLSLSPKLTIDIGFIGNFGLSVYDRKNSLSLSTGENYK---104  
666 GHLEOKSPASPNLSTFTW---DKTDA-----YQRYVGLSDAVVSV--GEYEYTCP 713  
105 --VIETDKTVLQOKLID--NL---RPEK---DLKENCYRIHK--SGDIEVLTG-----147  
714 SLIWEKRTALLOFELDPNLSLGGMSLDKHHILNVKSG---ILHKGTGENQFLTQOPALI 770  
148 ---FNNAFDLKYV-----KLLNPAGHAIIYIDMFEATOPRLNRIYDDLDGDHDP 195  
771 TSIWNGRRRSISCPSCNGLAEGNKLALPALAVGIDGSL-----810  
196 LILNLEYOGLITLILFPGQKEGRTETELFLNRQL---NSIHNSIGENMLTWSFGYT 251  
811 ---YGDENYIRIRIPSRNVTSILELR--NKEFKSNNPAHKXYLA--VDPVSGSLYVS 862  
252 PIGKNGILQWITMTAPGLKETVYVSNNOGHHFPOSNALPYLPYVTLAKQVPGAGOP 311  
863 DTNRRRIY---RVKSLISGTRKDLAGNSEVVAAG---TGEQCLPF---DEARCGPQK 908  
312 AIOAEYSYSHNVYGGSGNGIMN-----NKLNLVGLMTEYVYGGTESRRYK---DK 360  
909 AIDA--TMSPRGLAVDKNGLMYFVDATMIRKVDQ--NGIISTL--LGSNDLTAVRPLSCDS 964  
361 EGHQOYRIE-----RTNNNHLLTSECKQONGYIQTETAVYALIGNPFOSQBP 412  
965 SMDVAQVLEMPDPLAVNPMDSLYVL-----ENNVLIRITENHOVSTIA---GRPMHC 1015  
413 QLPRTKETWRSADNSYRSEITETTFDESGNPLTFVIRKDKTKOKIISPSTMEYYPAGE 472  
1016 QVPGIDYSLSLALHSALESASALASHTGVLVYITETDEKKINRLQVTTNGEICLLAGA 1075  
473 VDNCEPPE-----PYGFTREYVKKIIQTPYDSEFKDP 503  
1076 ASDDCDKDVANCYSGDAAYATDAIILNPSLSLAVAPDG--TIIYADLGNIRIRAVSKNKP 1134  
504 --EKPEIORYSLIGSOSHVTLKIEBRYHSAQOLNLSL---FOVYTKDESGRLILKQEC 558  
1135 VLANAQVEASPGQELVFNADSIHQYTVSLVGEYLVNFTSTID--NDVTELI-----1188  
559 TKGENGKTYSVVHKFTYTKODDTLQOSH-----SITTHDNFTIHSQ-----600  
1189 --DNNGNSLKT-----RDSGSGPRHLIMPNDQIITLVGTNGGLKLVVSTQNLIELDM 1239  
601 VRSRYTGLFSDPTDYDIYVQMSYDKLGRLLTRTLNLSGT-----PYANTLTUYEYLANL 654  
1240 TYDNGTGLLATKSDTGTGTTFYDYHBSGLTNVTPGTVTSYLHREMEKSTIIDIDENSR 1299  
655 QDDNRPPVITTTDVG-----NOLRNEFD-----GAGRH-----684  
1300 DDD-----VTYITNLSSVEASTYVQDOVRNSYQLCNNGTGLVMYKANGKGISFSHSEPHYL 1354  
685 -----VSQC-----LKDSGDGDKFYTIHQYDEQGRNHTSTYSDYL 721

1355 AGTITPTIGKCNLSLPMENGANSIEWRLRQKQKGVIFGRKLAVHGNLLSI--DYD 1411  
722 TNGR--QOTBDKXHLMSKSYDNWQ-----IANTHWSY-----GVSEK 759  
1412 RNIRTEKIIDHDKKFLRIIYDQVGRFLMLPSSGLAAVNVSYFFENGRLAGLORGANSER 1471  
760 ITVDPIPLATKOLQSNNSNVQTEKEVTTYTPSQPQIQLTFPEBAGHLSCHTLTRDGM 819  
1472 TDD-----KGRIVSRMFADGK--VMSYSLDKSMVLLQSORQ-----1510  
820 RVRKETDAIGQCTIYOYDYNRVLIQTLPDGTIVNRKPAFSTDTLITDIRVNGISLGGQ 879  
1511 -----IFEXDSDRLAAVTMPS--VAHH--SMSTHTSIGYIR-----NIYNP 1548  
880 TFDGLSRLOSQDGRVMAVTTYSAGNDQCSYVITTDGQRIHQY--QPELDAVYLQVAS 937  
1549 PERNASVIFDPSYDGGIL-----KTSFLGTGRQVFPYKGLSKLSLSEIVD--S 1594  
938 NEITQPSYNPVTGAL--LKAVALGOSLPRIYPSGRLKKNENIDMKQMSYLTWLRGLN 995  
1595 TAVT--FGYDETTGVLMVNLQSGCFSCIRYKIGPL-----VDKQIYRSEBGMVN 1645  
996 GYTDLGTLOKISRDTHGRVTOIKDSIKTTL-----NYDDLN--RHIGS-----QVT 1041  
1646 ARPDYT-----YHDSFRIASIKPVLSETPPLVDLYRYDBIGKVEHFGKFGVLYYDIN 1699  
1042 DLATGHMLTTTVPD--GLNREIGRLCDSGHTLDOIQSWLMTQOLANRIVKLN--GV 1096  
1700 QITTTAVMTLSKIPDTHGRVKEVOYEMFRSLMTMTVQYWSM-----GHVIRELKLGP 1753  
1097 LQRTBOQSYDSRNRLNQYKCDGA-----ECPDXYGSHSVTONFTYDIYGNITACHTT 1149  
1754 YANTTKTYYD-----YDSDGQLQSAVAVNDRFTMR-----SYDLNGNL-----1791  
1150 PADGEDHATFKRANPTDPCQLTEVHHHPDMEDNIRLKTDKKGRVNI-----TNGK 1203  
1792 -----HLNPNSSVRL-----MP-----LRVDLDRITRLDVOYKIIDDG 1827  
1204 -----NTNFTYDTLGL-----QNGQGSVYGVDPRLNRVLSOKTDPDLCELYYRETMV 1252  
1828 YLCQSGSDIFEVYNSKGLTALPAVNAKSGMSVOYRIDGGRASAYKTN--LGHILQYFYDHL 1886  
1253 NEVR-----NGEMIRLL--RTGETIIAQBPASKVLTGTDSQOS--VILTSDKQNTS 1300  
1887 NPRTITHVNHMSSEISLYYDLOGHILFAMESSGEYVYUASDNTGPRPLAVFSINGMLIK 1946  
1301 QEAYSAYGK--HKSTANDASILGNGERADPVSCVTHLGNGYASYDPTLMKRFHTPD-----1354  
1947 QLOQTAYGEIYYDSNPFCQWVIGFHGGLYDPLTGLVHFTQ--RDYDVLAGKMTSPDYTMW 2004  
1355 -----SLSPFGAGGINSYCIQDPINRS--DPSGHL-----SMQWMTGI-----1392  
2005 KNGKKEPAPF-----NLIMFSSNNPLSSELDKMYIVYDVKSMVLMFQFOLSNIIPGPPRA 2059  
1393 -----GNGIAGLLITATGMAIAAG-----IAAALA-----1421  
2060 KMVFPPPYELSESQASBENGLITGVQOTTERHNQAFMALEGGYITKXHLASIEKAGHW 2119  
1422 -SISTTALAAGAL-----SYTSDITSIVSGALDAPKASSILG-----1459  
2120 FATTPPIIGKIMAIKEGRVTTGVSSIAS--ED--SRKASVYLNNAUYLDKMYYSIEBK 2175  
1460 -----VWSMGMAAGLA-----ESA-----IKGSKLATIHGAFAEDGE 1493  
2176 DTHYFVFKIGSAGDPLVLTGTTIGKAVLESQVNTVSOPTLLVNGRTRAFNI--EQQY 2231  
1494 NALIKSTSESRIRIKMGT--RSLDREIVNRBEGQYIKDHSRGYTDNFMKGQOAILVHGDK 1552  
2232 STLLS-----IKYGLTPTLDEKAR-----VLQDARGLAGTAMAKQOQKARBD--R 2278  
1553 DGLYHTBGNK 1563

Db 2279 EGSRLMTEGEX 2289

RESULT 8

US-10-453-372-114

Sequence 114, Application US/10453372

Publication No. US2006000323A1

GENERAL INFORMATION:

APPLICANT: Alabrook, et al.

TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD

FILE REFERENCE: 21402-589 A

CURRENT FILING DATE: US/10/453,372

CURRENT FILING DATE: 2003-06-03

PRIOR APPLICATION NUMBER: 09/789390

PRIOR FILING DATE: 2001-02-23

PRIOR APPLICATION NUMBER: 60/185967

PRIOR FILING DATE: 2000-03-01

PRIOR APPLICATION NUMBER: 09/823187

PRIOR FILING DATE: 2001-03-29

PRIOR APPLICATION NUMBER: 60/195792

PRIOR FILING DATE: 2000-03-10

PRIOR APPLICATION NUMBER: 09/839446

PRIOR FILING DATE: 2001-03-19

PRIOR APPLICATION NUMBER: 60/199476

PRIOR FILING DATE: 2000-03-25

PRIOR APPLICATION NUMBER: 09/863776

PRIOR FILING DATE: 2001-05-23

PRIOR APPLICATION NUMBER: 60/208263

PRIOR FILING DATE: 2000-05-31

PRIOR APPLICATION NUMBER: 09/933398

PRIOR FILING DATE: 2001-08-24

PRIOR APPLICATION NUMBER: 60/227800

PRIOR FILING DATE: 2000-08-25

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 1609

SOFTWARE: Cureseq1st version 0.1

SEQ ID NO 114

LENGTH: 2662

TYPE: PRT

ORGANISM: Homo sapiens

US-10-453-372-114

Query Match 3.1%, Score 272.5; DB 6; Length 2662;

Best Local Similarity 19.5%; Pred. No. 2.8e-09;

Matches 365; Conservative 251; Mismatches 652; Indels 603; Gaps 99;

49 GHIWVNG-NLGPPLPLTSLYSPLNKTDIGFQIGFNGFSLVYDRKNSLSSTGENYK-- 104

995 GHLPQKSFQASPNLASFTIW---DKTDA-----YQGRVYGLSDAVSV--GPEYETCP 1042

105 --VIETKTYKLOOKKD--NL---RBEK---DLKENCRIIHK--SGDIEVLTG----- 147

1043 SLIWMKRTLLQGFELDPNSLQMSLDKHIIILNVKSG---ILHKGTGENQFLTQCPALI 1099

148 -----FNNAPDLKVP-----KKLLNPAGHAIYIDMFEATQPRMLNRIYDLDGDIP 195

1100 TSTMGNRRRSISCPSCNGLAEGNKLAPALAVIGDSL----- 1139

196 LNLMEYQGLIKTILTPPGQKEGYRTELRLNROL-----NSIHFSLGENPLTWSFGYT 251

1140 -----YGDFFVYIRIFPSRNVTSILELR--NKEFKSNNPAHKYULA--VDPVSGSLYVS 1191

252 PIGKNGILGOWITSMTPGGLKETVNTSNNQGHNFQOSANLPVLPYVTLTKQVGAQOP 311

1192 DTMSRRRY-----RVKSLGTKDLAGNSEVAVG-----TGEQCLPF---DEARCGDGXK 1237

312 AIOLEYVTSHTNYVGGSGIWN-----NKLDPNLGLMTEVYVGSSTESRRYK-----DK 360

1238 AIDA--TLNBPBGAVKNGKLMYFVDATMIRKVDQ--NGIISTL--LGSNDLTAVARPLESCDS 1293

361 EGHDIQVIRI-----RTYNNYHLLTSECKQONGYIQTETAYVAIIIGHNPDQSPQF 412

1294 SMDVAQVRLEWPTDLAVNPMDSNLYVL-----ENNVIILRTITENQVSIIA-----GRPMHC 1344

QY 413 QLPKTKETWRASADNSYRSEITETTPDESQNPILTKVLIKDKTKQKIIISPSTHMEYYPAGE 472

DB 1345 QVPGIDVSLSKLAHSHLESASALAIASHTGVLTYTETDEKKNLRLQVYTNGLICLLAGA 1404

QY 473 VDNCPPE-----PYGFRFVKKIIQTPYDESPKDDP 503

DB 1405 ASDCDKDNVNCYSGDDAVANDALINSSSLAVABDC--TIIYADIGNIRIRAVSKKP 1463

QY 504 --EKPIQYRSLIGSGHVTLKIEERYYSATOLLNSTL---FOYNTDKSELGRLLKOTEC 558

DB 1464 VLNAFNQYEAASPEQGLYVFNADGIHQYVSLVYGEVLYNFTYSTD--NDVTELI----- 1517

QY 559 TKGENGTYSVAKFTTKKODDTLQOSH-----SITTDNFTIRSQ----- 600

DB 1518 --DNNGNSLKI-----RDSGGMPRHLLMPDNIITLLTVGNGKLKVVSTONLEGLM 1568

QY 601 VRSRYGRLPSPDPTKDIYVQMSYDKRGLLTTRTNGST-----PYANTLTVOYELNLT 654

DB 1569 TYDGNITGLATKSDDEGTWTFPYDYDHEGRLLTNVTRPGVYVSLHREMEKSIITDIENSNR 1628

QY 655 QDDNRPPPVITTTDVNG-----NQLRNEPD-----GAGRH----- 684

DB 1629 DDD-----VTVITNLSEVSAVTVVQPVQVRSYQLCNNGTLRVVYANGMGISFHSBPHVL 1683

QY 685 -----VSQC-----LKSDGDKTYTHTHQYDQGRHHTSYGDL 721

DB 1684 AGTTPPTIGRCNLSLPENGNLSIEMRLRKEQIKGK--VTFGRRLVHGRLLSI--DYD 1740

QY 722 TNGR--QOTDDPKVHLSMSKSYDMWQ-----IANTMSY-----GVSEK 759

DB 1741 RNIRTEKIYDHKFTKRIITIIDYQGRFPLPSSGLAANVSYFENGRLAGLQGMASER 1800

QY 760 IYVDPITLTATKQLQSNNSNVQGEVYTYTPSQOPIQITLFBAGHLOSGCHTLTRDGWD 819

DB 1801 TDID-----KQRIYSRMFADGK--VMSYSYLDKMSVLLQSQORQ----- 1839

QY 820 RVKRETDALQCTIYQVDNANRYQITLPDGTIVNRKAPSPDTLITDIRVNGISLQOQ 879

DB 1840 -----IFEDSSDRLLAVTMPs---VAAR--SMSTHTSISGYIR---NTYNP 1877

QY 880 TFDGLSLRLOSQDGRVWATYASAGNDQCPSTVITPPGQPIHYQY--QPELDVAIVLAS 937

DB 1878 PESNASVIFPYSDDGRL-----KTSILGGRQVFPKYTGSLSLSELYVD--S 1923

QY 938 NEITQOFSYNPVYGAU--LKAVALGOSLPDIYPSGKLKMNENIDMKMSYMLTARGLEN 995

DB 1924 TAVT--FGYDETTGVLMKVMNLQSGGFSCTIRYKRIQPL-----VDKQIYRFBEGAMVN 1974

QY 996 GYTDLTGCTIOKISRDTHGRVTOIKDSIKTTL-----NYDDLN--RHIGS-----QVT 1041

DB 1975 ARPDYT-----YHDSFRIASIKPVISEPLPYDLRYBIEISGKVEFKFGVLYYUDIN 2028

QY 1042 DLATGMLLTTPVDP--GLNRBIGRKLCDSSGHTLIDIQSMLKTOQLANRIVKLN---GV 1096

DB 2029 QIITTAVMILSKFIDTGRKKEVOYEMFNSLMATMTVQYQSM-----GVYIKRELKGP 2062

QY 1097 LQRTQSYDNRRLNQYKCDGA-----ECPDKYSGHSIVTQNFYDIYGNITACHTT 1149

DB 2083 YANTTKTYD-----YDGDQLOQSAVANVDRPTMR-----SYDLNGLN----- 2120

QY 1150 PADGEDHATFKANPDPQOLTEVNAHTHPDMEDNIRLTKDKAGRYVNI-----TNNKG 1203

DB 2121 -----HLNPNNSVRL-----MP-----LRYDLDRITRLADVOYKIDDDG 2156

QY 1204 -----NTENFTYDTLGL-----ONGQGSVYGYDPLNRLVSOQTPDLDELYYRETMV 1252

DB 2157 YLCQSGSDIFENSKGLTAAVYKASGMSVOYRVDGGRASAKTN--LGHNLQYFYSDLH 2215

QY 1253 NEVR-----NGEMIRLL--RTGETTIAQORASKYLTLGTDSQOS--VILTSDKONS 1300

DB 2216 NPRTITHVNHNSSEITSLYYDQLGHLFAMESSSGEYVAVASDNTGTPRLAVFSINGIMIK 2275

Qy 1301 QEAYSAYGK--HKSTANDASILGNGERADPVSGVTHLNGRSGYDPTLMKFTPTD----- 1354  
Db 2276 QLOQTAYEIIYYDNDPQWVIGFRGGLYDPLTKLVHFTQ--RDYDVLAGRWTSPTDTMW 2333  
Qy 1355 -----SLSPGAGGINPYSYCLGDPINRS-DPSGHL-----SWQAWTGI----- 1392  
Db 2334 KNGVEKPAFP-----NLTYMFKSNNDPLSEELDKKNVYTVDKSMLVWFQGLSNIIIPGFPRA 2388  
Qy 1393 -----GNGIAGLLLTITANGMAIAAG-----IAAIA----- 1421  
Db 2389 KMVYVPPPELYESESQASENGQLITGVQOTTERHNOAFMALEGQVITTKQLHASIREKAGHW 2448  
Qy 1422 -STSTTALAFAAL-----SVTSDITSVSGALEDASPASITG----- 1459  
Db 2449 FATTPITIGKIMPAIKGRVTTGVSSIAS--ED-SRKVASVLNNAYLDMKHYSTIEGK 2504  
Qy 1460 ---WVSMGGAAGLA-----ESA-----IKGQTKLATHLGAPAEDEG 1493  
Db 2505 DTHYFVKIGSADGDLVLTGTTIGRKVLSEGVNVTVSQPTLLVNGRTREFINI-----EFQY 2560  
Qy 1494 NALKSTSESRIRIKGVT--RSLDREIVNNEGQVTKHDSRGVTDNFMKGKGOAILVHGDK 1552  
Db 2561 STLISL-----IRYGLPTDLDEKAR-----VLDAQRPALGTAMAKQOQARDS-R 2607  
Qy 1553 DGPLYHTEGKN 1563  
Db 2608 EGSRLWTEGK 2618

RESULT 9  
US-10-453-372-148  
; Sequence 148, Application US/10453372  
; Publication No. US2006000323A1  
; GENERAL INFORMATION:  
; APPLICANT: Alisbrook, et al.  
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD  
; FILE REFERENCE: 21402-589 A  
; CURRENT APPLICATION NUMBER: US/10/453,372  
; CURRENT FILING DATE: 2003-06-03  
; PRIOR APPLICATION NUMBER: 09/789390  
; PRIOR FILING DATE: 2001-02-23  
; PRIOR APPLICATION NUMBER: 60/185967  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: 09/823187  
; PRIOR FILING DATE: 2001-03-29  
; PRIOR APPLICATION NUMBER: 60/195792  
; PRIOR FILING DATE: 2000-03-10  
; PRIOR APPLICATION NUMBER: 09/839446  
; PRIOR FILING DATE: 2001-03-19  
; PRIOR APPLICATION NUMBER: 60/199476  
; PRIOR FILING DATE: 2000-03-25  
; PRIOR APPLICATION NUMBER: 09/863776  
; PRIOR FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: 60/208263  
; PRIOR FILING DATE: 2000-05-31  
; PRIOR APPLICATION NUMBER: 09/939398  
; PRIOR FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: 60/227800  
; PRIOR FILING DATE: 2000-08-25  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 1609  
; SOFTWARE: CuroSeqList version 0.1  
; SEQ ID NO 148  
; LENGTH: 2724  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-453-372-148

Query Match 3.1%; Score 272.5; DB 6; Length 2724;  
Best Local Similarity 19.5%; Pred. No. 2.9e-09;  
Matches 365; Conservative 251; Mismatches 652; Indels 603; Gaps 99;

Qy 49 GHIWNG-NLGPPLTLTSLYSPLNTDIDIGFIGIGFNGSLVYDRKNSLSTGENYK--- 104

Db 1057 GHLEKSFQASPNLASFIFW---DKTDA-----YQARYGLSDAVAVS--GREYETCP 1104  
Qy 105 --VLETKYTKLQOKLD--NL--REK-----DKENCRIIHK--SGDLEVLG----- 147  
Db 1105 SLIEMKRYTLALQGFELDPNSLKGWSLDKHIILNVSG--LHKGTGENQFLOQAPAI 1161  
Qy 148 -----FNNNAFDLKP-----KCLNPAGHAIIYIDMNEAQPRLNRIYDLDGDHP 195  
Db 1162 TSTMGKRRRSISGSPGNGLAEGNKLAFLVALAVGIDSL----- 1201  
Qy 196 LMLRYGLIKITLITLTPQKEGYRTELRLNQL-----NSIHFSLGENEPLTWSFGYT 251  
Db 1202 -----YGDENVYIRIRIPSHVTSIETLR--NKEFKSNPAHKXYLA--VDPASGSLYVS 1253  
Qy 252 PIGKNGILGQWISMTAPGLKETVNSNNQGHFQSANLPLPYVTLMKQVPGAGP 311  
Db 1254 DYSRRRIY-----RVKLSGTKDLAGSEVYAG-----TGEQCLP--DEARCGQGK 1299  
Qy 312 AIOAEYSYTHNVYGGSGNGIWN-----NKLNDLGLMTEYVYSGTESRRYK-----DK 360  
Db 1300 AIDA--TLMSPRGAVDKGLMFLVDTMTIRKVDQ--NGIISTL--LGSNDLFLAVRPLGDS 1355  
Qy 361 EGHDOIVRIE-----RTYNNYHLITSECKQONGYIQTETAYVAILGHNFDSPSQF 412  
Db 1356 SMDVAQVRLPPTDLAVNPMDNSLYVL-----ENNVLIRITENHQVSIIA-----GRPMHC 1406  
Qy 413 QLPKTKETWRSADNSRYSEITETTPDESGNPLRVKDKKTKXILSPSTMEYEPAGE 472  
Db 1407 QVPGIDVSLSLALHLSALBSASALAIASHTGVLTYTETDEKKINLRQVYTNGBICLLAGA 1466  
Qy 473 VDNCPPE-----PYGFTRFYKXIIQTPYDEFPKDP 503  
Db 1467 ASDCDKNDVNCYSGDDAAYTDALINSPSLAVADG--TIYADLGNIRIRAVSKNKP 1525  
Qy 504 --EKFIORYSLIGSQSHVTLKIEBRYSATQLINSTL--FOYNTDKSELGLLKQTEC 558  
Db 1526 VLNAPNGYEAAAPBQGLYVFNADGIHQYVSLVTEGYLVYFTSTD--NDVTBLI----- 1579  
Qy 559 TKGNGKTVSGVNHKFTYTKQDDTLQOSH-----SITHDNFTIHSQ----- 600  
Db 1580 --DNNGNSLKI-----RDSGMPRHLMLPDNOIITLTVGTNGGLKLVSTQWLEGLM 1630  
Qy 601 VRSRYTRLSPDYDTKQIVTQMSYDKLGRLLTRTINSGT-----PYANTLLTYDELANL 654  
Db 1631 TYDENTGLATKSPDETQMTTFYVDHDEGRLLNVTRPGVYVSLRBEKSTIIDIENSNR 1690  
Qy 655 QDDNRPPVITTTDVNG-----NQLRNEPD-----GAGRH----- 684  
Db 1691 DDD-----VTVITNLSVEASVTVVQDVNRNSYOLCNGGTLRVWYANGMGISFISBPHTL 1745  
Qy 685 -----VSQC-----LKSDGQKFTYTHITQOYDEGRRHHTSYSDYL 721  
Db 1746 AGTTPTTIGRCNLSLPMENGINSIEMRLRKEQIKGK--VTFGRKLVRHGRMLLSI--DYD 1802  
Qy 722 TNGR--QOQDPKVAHLSMSKSYDNMGQ-----IANTHWSY-----GVSEK 759  
Db 1803 RNITEKYIDHRKFTLRIIYDQGRPLMPLSSGLAIVNVSYFENGRLAGLGAGMSER 1862  
Qy 760 ITVDPITLTATKQLOSNNSNVQTEKEVTTYTPSQOPIQITLPDEAGHLQSCHTLTRDGD 819  
Db 1863 TDID-----KQGRIVSRMFADGK--VWSYSYLDKSWTLAQSORQY----- 1901  
Qy 820 RVRKETDALIGCTIYOYDNNVRVQITLPDQTIYNRKAPRSTDTLTDIRVNGISLQO 879  
Db 1902 -----LFEYDSSRLAIVTPS--VAARH--SKSTHISYIYR--NITNP 1939  
Qy 880 TFDGLSLRLOSGQSGRAWYATYSAGNDQCESTVITPPQGFTHQY--QPELDAAVLQVAS 937  
Db 1940 PESNASYVIFDYSDDGRL-----KTSFLGTRGQVFKYKCKSKLSSEIYYD--S 1985  
Qy 938 NEITQOQPSYNPVTGAL--LKAIVAGQSLTPIIYPSGRKLMENINDMKMSYLMWTIRGLEN 995



```

Db 1986 TAVT--FGYDETTGVLKVVNLQSGGFSCTIRYRKIGPL-----VDKQIYRFBEGMWN 2036
Qy 996 GYDLDLTCTQKISDTHGRVTOIWDSSIKTL-----NYDDL--RHIGS-----QVT 1041
Db 2037 ARPDYT-----YHDSFRIASIPVISETPPLVDLYRYEISGKVEHFGFVYIYDIN 2090
Qy 1042 DLATGMLTTTVEPD--GLNREIGRKCDSSGHTLDIQSGSLKQOLANRIVKLN---GV 1096
Db 2091 QIITTAVMTISKMDTHGRVKEVYEMFRSLMTMYTQYDSM-----GRVIRKELGCP 2144
Qy 1097 LQRTQYSYDSRNLNQCDDA-----ECPTDKYGHISVTQNFYDIYGNITACHTT 1149
Db 2145 YANTTKTYD-----YDGGQQLQSVAVNDRPFTMX-----SYDLNGLN----- 2182
Qy 1150 FADCTEHAFTKPAFPDPCQLTGVHTHTHPMPNINLKTDKAGRVNI-----TQNG 1203
Db 2183 -----HLNPNQSVRL-----MP-----LRVDLRDRIRLGDVQYKIDDDG 2218
Qy 1204 -----NTENFTYDTLGRL-----ONGQSVYGYDPLNRLVSOXTDLDLCELYYRETMLV 1252
Db 2219 YLCORGSDFEYNSKGLLTRAYNKASGWSVOYRVDGGRASVYTN--LGHHLQYFYSDIH 2277
Qy 1253 NEVR-----NGEMIRL--RTGETTIAQORASKYLITGDSQOS--VILTSKONLS 1300
Db 2278 NPTRTITHVNHNSNEITSLVYDLOGHLFAMESSGEERYVASDNTGTPPLAVFSINGLMIK 2337
Qy 1301 QEAYSANGK--HKSTANDASLIGNERADPVSGVTLGNGYSYDPTLMKRFHTPD--- 1354
Db 2338 QLOVTAAGEIYYDSNPFOWMIGFHGGLYDPLTCLVHFTQ--RDYDLAAGKMTSPDYTMW 2395
Qy 1355 -----SLSPFGAGGINFYSCYCLGDPINRS--DPSGHL-----SMQWMTGI----- 1392
Db 2396 KAVGKEAPF-----NLVYMKSNPNLSSELDLKNYIVDVKSWLVMPFGOLSNITPGFPRA 2450
Qy 1393 -----GWSIAGLLLTATGMAIAAG-----IAAIA----- 1421
Db 2451 KMYFVPPPEYELSESQASENGQLTGVQOTTERHQAFVALLGCVITTKKLASIEKAGHW 2510
Qy 1422 -SSTTALFAL-----SVTSDITISVGALEADAPKASIIIG----- 1459
Db 2511 FATTPTPIIGKIMAIKEGRVTGVSIIAS---ED-SRKVASVLMNAYVLDKMYHSIEGK 2566
Qy 1460 -----WVSMGMAAGLA-----ESA-----IKGTLATHLGAFAEDGE 1493
Db 2567 DTHYFVKGASAGDLVLTGTTIGRKVLDESQVNTVSOPTLLVNGKTRFTNVI---EFQY 2622
Qy 1494 NALLKSTSESSRIKMGVT--RSLDREIVRNEBGVYIKDHSRGYTDNFMGKGEQALVHGDK 1552
Db 2623 STILLS-----IRYGLTPTLDEKAR-----VLDOAQBALGTAMAKEQOKARQD-R 2669
Qy 1553 DGPLYHTEGUK 1563
Db 2670 EGSRLWTEGER 2680

```

# RESULT 10

```

US-10-453-372-136
; Sequence 136, Application US/10453372
; Publication No. US20060003323A1
; GENERAL INFORMATION:
; APPLICANT: Alsbjorn, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-589 A
; CURRENT APPLICATION NUMBER: US/10/453,372
; PRIOR APPLICATION NUMBER: 09/789390
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185967
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/823187
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195792
; PRIOR FILING DATE: 2000-03-10

```

```

; PRIOR APPLICATION NUMBER: 09/839446
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/199476
; PRIOR FILING DATE: 2000-03-25
; PRIOR APPLICATION NUMBER: 09/863776
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/208263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/939398
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/227800
; PRIOR FILING DATE: 2000-08-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1609
; SOFTWARE: CuiSeqList version 0.1
; SEQ ID NO 136
; LENGTH: 2733
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-453-372-136

```

```

Query Match 3.14; Score 272.5; DB 6; Length 2733;
Best Local Similarity 19.54; Pred. No. 3e-09;
Matches 365; Conservative 251; Mismatches 652; Indels 603; Gaps 99;

```

```

Qy 49 GHIVGNG-NLGPPLPLTSLSPKNTKIDIGFGRGLSYVDRKNSLSTGENYK--- 104
Db 1066 GHVFSQSPASPLNASTFW---DKTDA-----YGQRYGASDAVASV--GEVETCP 1113
Qy 105 --VLETDKTVKLOOKLD--NL--RPEK---DLKENCYRIHK--SGDIEVLTG--- 147
Db 1114 SLIMKERTLLGLFELDPNSLGMKSLDKHILNVSQ---LHKGGBNQPLTQPAII 1170
Qy 148 -----FNNAPDLKVP-----KCLNPAHAIIYDWNFEATQPLNRIYDLDGDIP 195
Db 1171 TSIMNGRRRSISCPSCGNLGAENKLLAPVALAVGIDSL----- 1210
Qy 196 LNLLEYQGLIKTLTLPQKGEGRTELRLNQL-----NSINFLSGENPLTWSGYT 251
Db 1211 ----YVGDENVYIRIRFPSSNVTSTILELR--NNEFKSNPAPKIKYYLA--VDPVSGSLYVS 1262
Qy 252 PIGNGILQWITSMTAPGLKETVYNSNNQGHHPQSANLPLPYVTLMKQVPGAGOP 311
Db 1263 DMSRRRY-----RVKSLSTGDLAGNSVYAG-----TGBQCLP---DEARCGDCK 1308
Qy 312 AIOAESYTSYSHNYVGGSGNIN-----NKLDNLGLMTEYNGSTESRRYK---DK 360
Db 1309 AIDA--TMSPRGIAVDKNGIAMTFVDATMIRKVDQ--NGIISTV--LGSNDLTAVAPLSCDS 1364
Qy 361 EGHDIYRIR-----RTYNNYHLLTSECKQONGIQOTETAYVAILIGHNDSQPSOF 412
Db 1365 SMDVAQVRLFWPTDLAVNPMDNSLYVL---BNNVILRLTTEHNOVSIIL---GRPMHC 1415
Qy 413 QLEPKTETWRSADNSYRSEITETTFDESQNPPLTKVYKDKTKOIIISPSTHWEYTPAGE 472
Db 1416 QVPGIDYSLSKAIIHSLBSASAIAISHTGVLYTTEDEKIRKLQVTTNGEILCLAGA 1415
Qy 473 VDNCPPE-----PYGFTRVFKIIQTYPYSEFPDP 503
Db 1476 ASDCDKNDVNCNCYGDDAYATDAIINSPSLAVAPDG--TIYIADIGNIRIRAVSNGKP 1534
Qy 504 --EKFTQYRSLGLSGSHVTLKIEBHYSATQULNSTL---FOYNDKSLGGLLKQTEC 556
Db 1535 VLNAFQYENASFGEBELVFNADGIHQYTVSLVGBEYLYNFTYSTD--NDVTBELI 1588
Qy 559 TKGENGKTVSVHKFTYTKDDTLQOSH-----SITTHDNFTIIRSQ----- 600
Db 1589 --DNNGNSLKI-----RDSGGMRRHLLMPDNQIITLVGVNGGLKXVSTQNLBGLM 1639
Qy 601 VRSRYTGRFLSPDITDKOIVTQMSYDKLGRLLTFLNSGT-----PYANTLYVDEIANTL 654
Db 1640 TYDNGTGLATKSDDEGWTTFYDYDHBGRLLNTRPLGVVTSIAREMEKSIITIDIENSNR 1699

```

QY	655	QODNRPPVITTTDNG-----NOLNRBFD-----GAGR-----	684
Dd	1700	DDD-----VYITNLISVEASTYVQDQKRSYQLCNGNGLRWYANGMGISFHSBEPVL	1754
QY	685	-----VSOC-----LKSDGDGKFYTIHTQOYDEQGRHHTSYSDYL	721
Dd	1755	AGTTPPTIGRCNLSLPMENGINSIEWRLRKEQIKG-VTIFGRKLRVHGRMLST--DYD	1811
QY	722	TNGR-QQIDPDKVHLMSKSYDNMQ-----IANTHVS-----GVSEK	759
Dd	1812	RNRKEKLYDHRKFTLRITTYDQGRPLMLPSSGLAVNVSYPFNGRLAGLORGANSER	1871
QY	760	ITVDEPITLTATKQLOLSNNTNVOGKEVYTTYPSQQPIOTITFDEAGHLQSCHTLTRDGD	819
Dd	1872	TDID-----KQGRIVSRMFADGK-VMSYSYLDKSMVLLQSQOY-----	1910
QY	820	RVRKETDLAGCCTIYQYDNRNVRVIOITLPDGIYNRKAPSTDTLTDIRVNGISLQO	879
Dd	1911	-----IFEDSSDRLLAVTMS--VARH--SMSTHISIDYIR--NINP	1948
QY	880	TFDGLSRLTQSDODGRVWAAYTYSAGNDQCPSTVITPDDQCFHYOY--QPELDAVLQVAS	937
Dd	1949	PESNASVIFDYSDDRIL-----KTSFLGTGRQVPRKYKGSLSKSEIYD--S	1994
QY	938	NEITQOFSYNPYTGA--LKAVAEQSILPTIYPSGRLEKMENTINMKMSSYLMTLRLEN	995
Dd	1995	TAVT--FGVDPTTGVLLKWNVLQSGFSCITIRKRKIGPL-----VDQIYRFSSEGMVN	2045
QY	996	GVTDLTGITQKISRPTHGRKVQIKOSLKTLL-----NYDDLN--RHIGS-----QVT	1041
Dd	2046	ARFDTT-----YHDSFRIASIKYVISETLPVULYAYDEISGVNEHFGKFGVYIYDIN	2099
QY	1042	DLATGHMLTTVEFD--GLNREIGRKLDSGHTLIDQOSWLKTQOLANRIVKLN--GV	1096
Dd	2100	QITTAVMVTLKGFHTGRKEVQEXEMFRSLMYMTVQYDSM-----GRAVYKRLKGP	2153
QY	1097	LQRTQOYSYDSDNRNLNOYKCOGA-----ECPDJKGHSIYQNGFYDIYGINTACHTT	1149
Dd	2154	YANTTKYTYD-----YDGGQLOLSVANNDKPTWRY-----SYDLNGNL-----	2191
QY	1150	FADGTEDHATEFANPPTPCOLTVEYHHTHPDMDIRLKYDKAGRVINI-----TDNHG	1203
Dd	2192	-----HLNPNGSVRL-----MP-----LRYDLRDRITRLSDVOYKXIDDDG	2227
QY	1204	-----NTENFYDTLGRL-----ONGQSVYGYDLPLRLVYSQKTDVTLDCSLYRETMV	1253
Dd	2228	YLCORGSDIFEXNSKGLLTRAYNRKASGMSVOYRXGVGRBRASUKYN-LGHHQYQYSDIH	2286
QY	1253	NEVR-----NGKMIRLL--RTGETIIAQORASKVLLTGDSDQOS--VILTSPKQMLS	1306
Dd	2287	NPTRITTHYVNSNSIETSLYDLOQHLRAMESSSEEEYVASDNTGCTPLAAPSINGLMIK	2346
QY	1301	QEAVSAYGK--HKSTANDASILGVNGERADPVSGVTHLGNGRYSYDPTLMRPHND	1354
Dd	2347	QLOQYRAYGEIYVSDNPDEQMVIGFNGGLYDPLTKLVHFTQ--RDYDVLAGRWTSBDYTMW	2404
QY	1355	-----SLSPFGAGGINPYSYCLGDILNS--DSGHL-----SMQANTGI--	1392
Dd	2405	KNVGKEPAPF-----NLTYFKSNPNPISSELDLKNVTVDKYSVLMVFGQLSNIRGPPRA	2455
QY	1393	-----GNGIAGILLTLIATGMAIAAAG-----IAAIA-----	1421
Dd	2460	KMYFVPPPELSEQASENGQOLITGVQOTTEHHNQAPALBEOQVITKKLHMSIRKKACHW	2519
QY	1422	-STSTTALAFCAL-----SVTSDTITSYSGALEDASPASTILG-----	1455
Dd	2520	FATYTPRIIGKGMFAIKEGRVTTGVSSIAS--ED-SRKVASVLNNAAYLDMKMHYSIEGK	2575
QY	1460	-----VWSMGMGAAALA-----ESA-----IKGGTKALTHGAPABOGE	1492
Dd	2576	DTHYVVKIGSADGDLVLTGTTIGRKVLBSGVANVWYSQPLLVNGTKRTPNTI--EYO	2631
QY	1494	NALLKSTSESSHIKMGVNT-RSLDRERI VNRNEGOVLIKDSRGVTDNFMKGEOAILVHDK	1552

```

Db      2632 STILLS-----IRGGLPDTIDEKAR-----VDDQARALSTAMALEQQAKARDG-R 2678
QY      1553 DGPLYTEGNK 1563
          :|||:
Db      2679 EGSRLMTEGEK 2689

RESULT 11
US-10-453-372-142
; Sequence 142, Application US/10453372
; Publication No. US20060003323A1
; GENERAL INFORMATION:
;   APPLICANT: Alsobrook, et al.
;   TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHODS
;   FILE REFERENCE: 21402-589 A
CURRENT APPLICATION NUMBER: US/10/453,372
CURRENT FILING DATE: 2003-06-03
PRIOR APPLICATION NUMBER: 09/789390
PRIOR FILING DATE: 2001-02-23
PRIOR APPLICATION NUMBER: 60/185967
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: 09/823187
PRIOR FILING DATE: 2001-03-29
PRIOR APPLICATION NUMBER: 60/195792
PRIOR FILING DATE: 2000-03-10
PRIOR APPLICATION NUMBER: 09/839446
PRIOR FILING DATE: 2001-03-19
PRIOR APPLICATION NUMBER: 60/199476
PRIOR FILING DATE: 2000-03-25
PRIOR APPLICATION NUMBER: 09/863776
PRIOR FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: 60/208263
PRIOR FILING DATE: 2000-05-31
PRIOR APPLICATION NUMBER: 09/939398
PRIOR FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: 60/227800
PRIOR FILING DATE: 2000-08-25
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 1609
SOFTWARE: CuroSeqList version 0.1
SEQ ID NO 142
LENGTH: 2733
TYPE: PRT
ORGANISM: Homo sapiens
US-10-453-372-142

Query Match           3.1%; Score 272.5; DB 6; Length 2733;
Best Local Similarity 19.5%; Pred. No. 3e-09;
Matches 365; Conservative 251; Mismatches 652; Indels 603; Gaps 99;

QY      49 CHIVANG-NLGPTLPILTSYSPLNKTDIGFGIFNGLSYDRKNLSLSTGENYK--- 104
          ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1066 GHLPKSPASPNLSTPIW---DKTDA-----GGORYGSDAVSV--GREYETCP 1113
          :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      105 --VIETDKVKLQOKKD--NL---RFEK---DLXENCRRIHK--SGDIEVLVG----- 147
          ::||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1114 SLIWEKRTALLQGFELDPNIGWSLDKHHLIAVSG---LIHKGGENQFLIQQPAIL 1170
          |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      148 ----FNNAFDLKVP-----KKLLNPAGHAIIYIDWNFEATQPLNRITYDDLGDHIP 195
          |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1171 TSIMONGRRRSISPCSCNGLAEGNKLLAPVALAVGIDGSL----- 1210
          |||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      196 LLNLEYOGLKITLILFPQOKEGRYTELAFLNQL----NSINFSIGNENPLTWSPGYT 251
          |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1211 ----YVDGFNYIRIRIPSRNTSILIELR--NEEFHSNNPAHKYYLA--VDPSGSLVS 1262
          |||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      252 PIKGNKILGOWTTMTAPGLKETVAVSYNNQGHFPOSANLPLPYVTLLMKQVPGAGQP 311
          |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1263 DTNSRRIT----RVKSLSGTDLADGNSEVVAG-----TGECCLP---DEARGDGSK 1308
          |||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      312 AIQAESTYSTSHNVYGGGSGGINN-----NKLDNIYGLMTEINYGSTESRRYK---DK 360
          |||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

Db 1309 AIDA--TLMSPRGIAVDKGLMVFVDAWTMRKVQD--NGIISTL--LGSNDLJAVAPLSGDS 1364  
 Qy 361 EGDHDIYRIR-----RTYNNHLLTSECKQNGYIQTTERAYVALIGHNPFOSQSO 412  
 Db 1365 SMDVAOYRLWMPDPLAVNPMDSLYVL-----ENNVLIRITEMQVSIIA-----GRPMHC 1415  
 Qy 413 QLPKTKETRSADNSYRSEITETTPDESQNPFLKVKDKKTKOKIISPTMWEYPAGE 472  
 Db 1416 QVGEIDYSLSKLAHSLAESASALASHITGVLYTTERDEKKINRLROVTTNGEICLLAGA 1475  
 Qy 473 VDNCPPE-----PYGFTFRVYKIIQTPYDSEFKDDP 503  
 Db 1476 ASDDCDKNVNCYCSGDADAYATDALINSPSLAVAPDQ--TIYADIGNIRIRAVSNKP 1534  
 Qy 504 --EKFIOYRSLIGSOSHVLKIEBRYSATQLANSTL--FOYNTKSELGRLKTEC 558  
 Db 1535 VLNAFNQYEAASPEQELVFNADGHOYTVSLVTEGYLYNFYSTD--NDVTELI----- 1588  
 Qy 559 TKGENGTYSVNHKFTYTKODDTLQOSH-----SITHTDNFTIHSQ----- 600  
 Db 1589 --DNNGSLKI-----RDSGCMFRLMLPNDQIITLVGTNGGLKLVSTONLEIGLM 1639  
 Qy 601 VRSRYTSLPSDITDKOIVTQMSYDKLGRLLTRTLNSGT-----PYANTLYDYELANTL 654  
 Db 1640 TYDQNTGLATKSDQETGWTTFYDYDHGRLTNVTRPGVTVSLHREMEKSTIIDENSNR 1699  
 Qy 655 QDDNRFPVYITTVVNG-----NOLNREF-----GAGRI----- 684  
 Db 1700 DDD-----VTVITNLSSEVSAVTVVQDOVNSYOLCNNGLRVVYANGMGISFSEBPHVL 1754  
 Qy 685 -----VSQC-----LKSDGDKFYTHITQOYDQSGHHTSYSDYL 721  
 Db 1755 AGITTPITGRCNISLPMENGLNLSIEMRLRKEQIKGK-VTIFGRKLRYHGNLISI--DYD 1811  
 Qy 722 TNGR--QCTDPDKVHLNMSKSYDMNGQ-----IANTHNSY-----GVSEK 759  
 Db 1812 RNIRTEKIYDHRKFTLRIRIYDQGRPLMLPSSGLAANVSYPFNRLAGLQKANSER 1871  
 Qy 760 ITYDPTILTLATKQLOSNNSNVQSKVYTTTPSQOPIQITLFDAGHLQSHLTTRDQMD 819  
 Db 1872 TDD-----KQORIVSRMFADK--VMSYSYLDKSNVLLQSGROQY----- 1910  
 Qy 820 RVRKETDAIGCCTIYOYDNVNRVIOITLPOGTIVNRKXAPSTDTLTDIRVNGISLQOQ 879  
 Db 1911 -----TFEYDSSRLAIVTMS--VAKH--SMKSTITSIGYR--NIYNP 1948  
 Qy 880 TFDGLSLTQSDQGRVWAAVYYSAGNDQCSPTVITPDQFPIHQY--QPELDDAVLQVAS 937  
 Db 1949 PESNASVIFPQYSDGRIL-----KTSPLGGRQVFFYXGKLSKSEIYVD--S 1994  
 Qy 938 NEITQOSYNPVTGAL--LKAVAEGOSLTPITYSGKLEKININDMKMGSLYTLRGLN 995  
 Db 1995 TAVV--FGYDETTVLKQVNLQSGFSCITIRYRIGPL-----VKKOYRFSBEGMVN 2045  
 Qy 996 GYTULTGTIGIKSHDTHGRVLTQIDSSIKTLL-----NYDDLN--RHIGS-----QVT 1041  
 Db 2046 ARFPYT-----YHNSFRILASIRPVISETPLPVLDLYRYDLSKVEHFGGVIYYIDIN 2099  
 Qy 1042 DLATGHNLLTTFVEPD--GLNREIGRKLDSGHTLIDIQOSWLTQOLANRIVKLN--GV 1096  
 Db 2100 QIITTAWMTLSKHFDTHGRIRKEVOYEMFRSLMYMTMTQOYBSM-----GRIKRELKGR 2153  
 Qy 1097 LQRTQOYSYOSRRLNOYKCDGA-----ECPTDKYGHSLVTQNFYUDDYIGNTLACHTT 1149  
 Db 2154 YANTTKTYD-----YDGGGOQASVAVNDRPMTRY-----SYDINGNL----- 2191  
 Qy 1150 FADQTEHATFKFANPTDPCOLTEVHTHPMDPNIRLKYDKAKGRVNI-----TDNHG 1203  
 Db 2192 -----HLNFGNSVRL-----MP-----LRIDLRLRITRLDQVYKIDDDG 2227  
 Qy 1204 -----NTENFTYDTLGLR-----ONGQGSVYGYDPLNRLVSQKTDPLDCELYURETMY 1252  
 Db 2228 YLCORGSDIFRYNSKGLLTRYAUNKASGMSVOYRYDVGRRASUKTN--LGNHLQYFYSDLN 2286

Qy 1253 NEVR-----NGEMIRLL--RTGETIINOGRASKVLITGTSQOS--VILTSKONLS 1300  
 Db 2287 NPIRITHVNHNSSEITSLTYDLOGLHLPAMESSGEERYVASDNTGTPLAVFSTINGMIK 2346  
 Qy 1301 QEAVSAVGK--HKSTANDASILGNGERADPVSGVTHLGNYSRYDPTLRFTPD----- 1354  
 Db 2347 QLOVTAIVGELIYDSNPDQFQVIFGHLGYPDLTKLVHFTQ--KDYDVLAGRMTSPDYTMW 2404  
 Qy 1355 -----SLSPFGAGINDYSYCLADPPIHRS--DPBGHL-----SQOAWGTI----- 1392  
 Db 2405 KNYGKEBAPF-----NLWFKSNPNPLSSELDLKNYVYDVGSMVMEFGFOLSNIIIGPPRA 2459  
 Qy 1393 -----GNGIAGLLTATGMAIAAGS-----IAATA----- 1421  
 Db 2460 KMFTVPPPYLSSQASENQULITTVQOQTERNNQAFMALEGOVITTKGLASIREKAGHW 2519  
 Qy 1422 -STSTTALAFGAL-----SVTSDITSIVGSALEDAAPKASILG----- 1459  
 Db 2520 FATTPTIIGKIMFALKEGVTGVSISAS--ED--SRKVASVILNNAVYLDKXHYISIEGK 2575  
 Qy 1460 -----WVSMGMAAGLA-----ESA-----IKGTGLATHLQAFADGB 1493  
 Db 2576 DTHYFVYIGSADDLVTIAGTTIGRKVLESQVNTVSQPTLLVNGRTFRFTNI--EFOY 2631  
 Qy 1494 NALLKSTSESSRIKMGVT--RSLDRBYRNEEGVYIYDHSRGYDNTFKGKEQALIVHGDK 1552  
 Db 2632 STLLS-----IRYGLTPTLDEBKAR-----VLDQARQALGTMAKEQOKARDG-R 2678  
 Qy 1553 DGFYHTEGKNK 1563  
 Db 2679 EGSRLWTEGSK 2689

RESULT 12  
 US-10-453-372-146  
 ; Sequence 146, Application US/10453372  
 ; Publication No. US20060003323A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Alabrook, et al.  
 ; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD:  
 ; FILE REFERENCE: 21402-589 A  
 ; CURRENT APPLICATION NUMBER: US/10/453,372  
 ; PRIOR APPLICATION NUMBER: 09/789390  
 ; PRIOR FILING DATE: 2001-02-23  
 ; PRIOR APPLICATION NUMBER: 60/185967  
 ; PRIOR FILING DATE: 2000-03-01  
 ; PRIOR APPLICATION NUMBER: 09/823187  
 ; PRIOR FILING DATE: 2001-03-29  
 ; PRIOR APPLICATION NUMBER: 60/195792  
 ; PRIOR FILING DATE: 2000-03-10  
 ; PRIOR APPLICATION NUMBER: 09/839446  
 ; PRIOR FILING DATE: 2001-03-19  
 ; PRIOR APPLICATION NUMBER: 60/199476  
 ; PRIOR FILING DATE: 2000-03-25  
 ; PRIOR APPLICATION NUMBER: 09/863776  
 ; PRIOR FILING DATE: 2001-05-23  
 ; PRIOR APPLICATION NUMBER: 60/208263  
 ; PRIOR FILING DATE: 2000-05-31  
 ; PRIOR APPLICATION NUMBER: 09/939398  
 ; PRIOR FILING DATE: 2001-08-24  
 ; PRIOR APPLICATION NUMBER: 60/227800  
 ; PRIOR FILING DATE: 2000-08-25  
 ; Remaining Prior Application data removed - See File Wrapper or PALM.  
 ; NUMBER OF SEQ ID NOS: 1609  
 ; SOFTWARE: CuraSeqList version 0.1  
 ; SEQ ID NO 146  
 ; LENGTH: 2733  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-10-453-372-146

Query Match 3.1%; Score 272.5; DB 6; Length 2733;  
 Best Local Similarity 19.5%; Pred. No. 3e-09;  
 Matches 365; Conservative 251; Mismatches 652; Indels 603; Gaps 99;

QY 49 GHVYNG-NIGPTPLPLSLYSPLKTDIGFSGIGNFGLSYDRKNSLSLSTGNRYK--- 104  
 DB 1066 GHLEFQKSFQASPNLSTPTW---DKTDA-----YQQRVYGLSDAVAVSV---GREYETCP 1113  
 QY 105 --VLETKTVAGLQOKKD--NL-----RPEK-----DLKNCYRIHK--SGDIEVLG----- 147  
 DB 1114 SLIMEKRTALLQGFELDPNLSGKSLDKHIIIVKSG--ILHKIGENGFPLTQQAII 1170  
 QY 148 ----FNNAFPLKVP-----KCLINPAGHAIIYIMNFEATQPLNRIYDDLGDHDP 195  
 DB 1171 TSIMNGRRRRIISCPSCNGLAEGNKLLAPVALAVIGDSL----- 1210  
 QY 196 LNLNLYOGLKITTLTPPGQKEGYRTERFLNRLQ---NSIHNPSLGNPLTWSTGYT 251  
 DB 1211 ----YVGDFFNYIRIIPSRNVTSLLELR--NKSEFKSNNPAHKYULA--VDPVSGSLYVS 1262  
 QY 252 PIGKNGILGQWITSMTPAGLKEETVNSNNQGHFPOSANLPLPYVTLMKQVPGAGOP 311  
 DB 1263 DTNSRRYI----RVKSLSTGKDLAGNSEVYAG-----TGEQCLPF---DEACSGQSK 1308  
 QY 312 AIOAEYSYTHNYVGGSGNGIMN-----NKLDNLYGLMTEYNGSTESRRYK---DK 360  
 DB 1309 AIDA--FLMSPRGIAVDKNGIMFYVDATMIRKQD--NGIISTL--LGSNDLPAVAPLSCDS 1364  
 QY 361 EGHQOIVRIE-----RTYNNHLLTSECKQONGYIQTETRYVALIGNPFSQSP 412  
 DB 1365 SMDVAQVRLWPRTDLAVNPMDSLYVL-----ENNVLIRITENHOVSIIA---GRPMHC 1415  
 QY 413 QLPKTKETWMSADNSYSEIETETTFDESGNPLFKVIKDKKTOKIISPSTMEYYPAGE 472  
 DB 1416 QVPEIDISLGLALHSALESASALAISHTVLVIYTEDKGIINLRQVTNGELCLAGA 1475  
 QY 473 VDNCPPE-----PYGFTFVYKIIQTPYDSEFKDP 503  
 DB 1476 ASDDCDKNDVANCYSGDDAVATDAIINSPSLAVAPDG--TIYADLGNIRIRAVSNKP 1534  
 QY 504 --EKPIQRYRLISSQSHVTLKIEBRHYSATQLANSTL---FOYNTDKSELGRLLKQTEC 558  
 DB 1535 VLNAFNQVEAASPEQELVYENADGIIHQYTVSLVGEVLYNFTYSTD--NDVTELI----- 1588  
 QY 559 TKGNGKTYSVVHKFTYKODDTLOQSH-----SITTDNFTHRSQ----- 600  
 DB 1589 --DNNGNSLKI-----RDSGSGPRHLLMPDNOIITLVGTNGKLVSTQMLEGLM 1639  
 QY 601 VRSRYTGLFSDDTKDIIVTQMSYDKLGRLLTTLNLSGT-----PYANTLVYDELANL 654  
 DB 1640 TYDGNLTGLATKSDDEGTTFPYDYDHEGRLNVTREPTGVVLSLHREMEKSIITIDIENSR 1699  
 QY 655 QODNRPPVITTTDVNG-----NOLRNEFD-----GAGR----- 684  
 DB 1700 DDD-----VTVITLMSVEASYTVVQDVNSYQLCNNGTLRVMYANGMGSIFSEPHVL 1754  
 QY 685 -----VSQC-----LKDSGDGKFTYIHTQOYDEQGRHHTSYSDYL 721  
 DB 1755 AGITPTTIGRCNISLPMENGLNSTEMRLKQIKQK--VTIFGRKLKRVGRMLSL--DYD 1811  
 QY 722 TNGR--QOTDPPKVLHLSMSKSYDNMGQ-----IANTHWSY-----GVSEK 759  
 DB 1812 RNIRTEKYIDHRRKFTLRITIVQVGRPLMLPSSGLAIVANSYFENGRLAGLQSGANSEK 1871  
 QY 760 ITVDPITLTATKQLOSNNNVOTGEVNTYTPSOQPIQITLPEDEAGHQSCHITLTDQWD 819  
 DB 1872 TDID-----KQGRIVSRMPADGK--VWSYSTLDLSMWLLQSQKQY----- 1910  
 QY 820 RVRKETDAIGCTIYQVNYNRNVIITLPDGTIVRKRAKAPSTDTLIDIRVNGISLQO 879  
 DB 1911 -----IFEYDSBRLAVTMP--VARH--SMSTHSTIGYR-----NINYP 1948  
 QY 880 TFDGLSRLTQSGDGRVMAVYTYASGNDQCPSTVITPPDQGFHYQY--QPELDVAVLQVAS 937

DB 1949 PESNASVIFPYSDDGHL-----KTSFLGTRQVFFYKXGSLSLSEIYVD--S 1994  
 QY 938 NEITQGFVNPVYVAGL--LKAVAEQSLPFIYPSGRLKMEINNDKMSGYLTIRGLEN 995  
 DB 1995 TAVY--FGDETTGLKMWVLOSQGSCTIRYKIGPL-----VDKQIYRSEBQVN 2045  
 QY 996 GYVTLGTITOKISRDTHGRVTOIKDSIKTTL-----NYDDLN--RHIGS-----QVT 1041  
 DB 2046 ARPDYT-----YHNSFRILASIKPVSIEPLVLDYRYDEISGKVEHFGKFGVIYVDIN 2099  
 QY 1042 DLATGHMLTTTVEFD--GLNREIGRLCDSSGHTLDDIQSWLKTQOLANIVKLN--GV 1096  
 DB 2100 QIITVAMTLSKHPDTHGRVKEVOYEFRLMVMYVQYOSM-----GVVIRKELKGP 2153  
 QY 1097 LQRTQSYSDSRRLNQKCDGA-----ECPTDKGHSIVTQNFYTYGNITACHTT 1149  
 DB 2154 YANTTKTYD-----YDGGQLOQSVAVNDRPTWRI-----SYDLGNL----- 2191  
 QY 1150 PADGTEDHATFKFANPTPCQLTEVHTHPDMPDNIRLKYDKAGRVINI-----TDNHG 1203  
 DB 2192 -----HLNPNGSVRL-----MP-----LRYDLDRITRLGDVQYKIDDDG 2227  
 QY 1204 -----NTENFTYTLGL-----QNGQGSVYGYDPLNRLVSOXTDTLDCELYRETMV 1252  
 DB 2228 YLCORGSDIPEYNSKGLTAYNKAQSGSVQYRYDVGRAASYKTN--LGHLOFYFSDLH 2286  
 QY 1253 NEVR-----NGEMIRLL--RTGETIIAQBASKVLLTGSTQOS--VILTSQONLS 1300  
 DB 2287 NPTRITVYHNSSEITSLYYDLOGLHFMESSEGEYVYASDNTGTPPLAVPSINGLMIX 2346  
 QY 1301 QEAYSAYGK--HKSTANDASILCYNGERADPVSGVTHLNGYRSYDPTLMRPHTPD--- 1354  
 DB 2347 QLOYTAYGELIYDQSNPDQVWVIFGHGLYDPLFLGLVHFTQ--RDYDLAGRMTSPDYTMW 2404  
 QY 1355 -----SLSPGAGINPYSCLDPPNRS--DPSGHL-----SQWATYGI----- 1392  
 DB 2405 KMWKEKPAFP-----MLWPKSNNPLSSELDKNYVYDVXSWVWFMFPQSLNIIIPGPRA 2459  
 QY 1393 -----GMGIAGLLTLATGGMALIAAG-----IAAIA----- 1421  
 DB 2460 KMTFVPPPEYLSQSQSENGQLITVQOTTERHQAMALEGVYITKGLASIEKAGHW 2519  
 QY 1422 -STSTLALFAGL-----SVTSDITSIVSGALEDAPKASIIIG----- 1459  
 DB 2520 FATTPTIIGKIMFALKEGAVTTGVSSIAS--ED--SRKVASVLNAAVYLDKMHYSIEGK 2575  
 QY 1460 -----WVSMGGAAGLA-----ESA-----IKGTLATTLGAPADGE 1493  
 DB 2576 DTHYFVYIGSADGDLVTLGTTIGRKVLSEGVNVTVSQPTLLVNGRTERRFTNI---EFQY 2631  
 QY 1494 NALLKSTSESRIKMGVT--RSLDEIYRNEGVYIKDSRGYTDNFMKGEOALIVHGDK 1552  
 DB 2632 STLLLS-----IRYGLTPTLDEKAR-----VLDQAORALGTMAAEQOQARQD--R 2678  
 QY 1553 DGFLYHTEGNK 1563  
 DB 2679 EGSRLMTEGER 2689

RESULT 13  
 US-10-453-372-150  
 ; Sequence 150, Application US/10453372  
 ; Publication No. US20060003323A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Alcobrook, et al.  
 ; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD  
 ; FILE REFERENCE: 21402-589 A  
 ; CURRENT APPLICATION NUMBER: US/10/453, 372  
 ; CURRENT FILING DATE: 2003-06-03  
 ; PRIOR APPLICATION NUMBER: 09/769390  
 ; PRIOR FILING DATE: 2001-02-23  
 ; PRIOR APPLICATION NUMBER: 60/185967

```

    1  PRIOR FILING DATE: 2000-03-01
    2  PRIOR APPLICATION NUMBER: 09/823187
    3  PRIOR FILING DATE: 2001-03-29
    4  PRIOR APPLICATION NUMBER: 60/195792
    5  PRIOR FILING DATE: 2000-03-10
    6  PRIOR APPLICATION NUMBER: 09/839446
    7  PRIOR FILING DATE: 2001-03-19
    8  PRIOR APPLICATION NUMBER: 60/199476
    9  PRIOR FILING DATE: 2000-03-25
    10 PRIOR APPLICATION NUMBER: 09/863776
    11 PRIOR FILING DATE: 2001-05-23
    12 PRIOR APPLICATION NUMBER: 60/208263
    13 PRIOR FILING DATE: 2000-05-31
    14 PRIOR APPLICATION NUMBER: 09/939398
    15 PRIOR FILING DATE: 2001-08-24
    16 PRIOR APPLICATION NUMBER: 60/227800
    17 PRIOR FILING DATE: 2000-08-25
    18 Remaining Prior Application data removed - See file Wrapper or PALM.
    19 NUMBER OF SEQ ID NOS: 1609
    20 SOFTWARE: ChraSeqList version 0.1
    21 SEQ ID NO: 150
    22 LENGTH: 2733
    23 TYPE: PRT
    24 ORGANISM: Homo sapiens
    25 US-10-453-372-150

    Query Match      3.1% Score 272.5; DB 6; Length 2733;
    Beat Local Similarity 19.5%; Pred. No. 3e-09;
    Matches 365; Conservative 251; Mismatches 652; Indels 603; Gaps 99;

    Oy 49 GHIYNG-NLQPTLP.LTISYPLMNTDIDGFGIFGSLVYDKNKSLSTGENYK-- 104
    Db 1066 GHLQKSFQASPNLASTPIW-----DKTDA-----YGRVYGLSVAVSV--GFEYECIP 1113
    Oy 105 --VIETDPTVKLQOKKLD--NL---RFEK---DLKENCYRLIHK--SGDIEVLTG----- 147
    Db 1114 SLIMKERTALLQGEELDPNSLGCNSLDKHIHLNWKSG---ILMKGTENQPTLQOPAI 1170
    Oy 148 ---FNNAFPLKVP-----KKLLNPAGHAIIYDNNFEATOPRLNRIYDDLDGHDIP 195
    Db 1171 TSIMNGRRRSISCPSCNGLABGNKTLAPVALAVGIDSL----- 1210
    Oy 196 LNLLEYQGLKITLILFPQOKRGRTTELRLNROL---NSIHNFSLGNENPLTWSFGYT 251
    Db 1211 ----YVDFNYIRIRIPFSRNVTSILER--NKEFKSHNNPAHKXYLLA--VDPVSGSLTVS 1262
    Oy 252 PIKKNGLIGOMITSMTAGGLKETVYVGNNOGHHPQSANLPLVRYTLTKQVPAGQRP 311
    Db 1263 DTNSRRY----RVKSLSGTDLAGNSEVVAG-----TSEQCIPF--DEARCGDGK 1308
    Oy 312 AIOAEVSYTHNVYGGSGNGIWN-----NKLDELVLGMLTEVYNGSTESRRYK----DK 360
    Db 1309 AIDA--TLMSPRGIAVDKNGMLFPVDMATMIRKVDQ--NIIISTL--LGSNDLTAIVRPLSCDS 1364
    Oy 361 EGHDOIVRIE-----RTYNNVHLTSECKQONGYIQTETAVAYAILIHNPDSPSQF 412
    Db 1365 SMDVAQVLEWPTDLVAPMNDNSLYVL-----ENNVIIRITENHQVSIIA---GRPMHC 1415
    Oy 413 QLPKTKETWRSADNSYSREIETTFEDSSGNPLTVIYIDKTKQKILISSTHMEYTPPAG 472
    Db 1416 QVPGIDVSLSKLAHLSALESASALAIHSHTGVLYTTEDEKKINRLRQVTTNGECLLAGA 1475
    Oy 473 VDNCPPE-----PYGTFRVFKIIQOTPYDSBFKDP 503
    Db 1476 ASDDCNDKDVANCNCYSGDDAATADAILNSPSSLVAAPQG--ITTYADLGNIRRAVSQKRP 1534
    Oy 504 --EKPIQIRYVLLIGSQSHVTLKIEBRHYSATQLNSTL--FOYNTDKSELGRLLIKQTEC 558
    Db 1535 VLNAFNOYEASPGQOELVFNADGIHQYTVSLVGEVLVNFYSTD--NDVTBLI----- 1588
    Oy 559 TKGBNGKTVSVVHKFTYTKODDTLOQSH-----STTHDNFIHHSQ----- 600
    Db 1589 --DNNGNLSKI-----RDSGGMFRLLMLPMDNQITLLTVGTINGGKLVSATONLEIGLM 1639

```

Qy	601	VR\$N\$YR\$R\$P\$D\$T\$K\$O\$I\$V\$Q\$M\$S\$Y\$D\$K\$G\$R\$L\$T\$R\$T\$U\$N\$G\$T\$-----P\$A\$N\$T\$T\$Y\$Y\$E\$A\$N\$L\$	654
D\$	1640	T\$Y\$D\$G\$N\$T\$G\$L\$A\$T\$K\$S\$E\$T\$G\$W\$T\$T\$F\$Y\$D\$H\$E\$R\$R\$L\$N\$V\$T\$P\$T\$G\$V\$T\$S\$L\$R\$E\$M\$E\$K\$S\$T\$T\$D\$E\$N\$S\$R\$	1699
Qy	655	Q\$D\$N\$R\$P\$P\$V\$T\$T\$D\$V\$N\$G\$-----N\$Q\$R\$N\$E\$P\$-----G\$A\$R\$H\$-----	684
D\$	1700	D\$D\$-----V\$V\$T\$T\$M\$L\$S\$E\$A\$S\$Y\$T\$V\$Q\$Q\$V\$R\$N\$S\$Y\$Q\$L\$C\$N\$N\$G\$T\$R\$V\$V\$Y\$A\$N\$Q\$M\$G\$S\$F\$H\$E\$P\$H\$V\$L\$	1754
Qy	685	-----V\$Q\$C\$-----L\$K\$D\$G\$G\$K\$Y\$T\$T\$H\$T\$Q\$Q\$Y\$D\$E\$G\$R\$H\$T\$S\$Y\$T\$L\$	721
D\$	1755	A\$G\$T\$T\$P\$T\$G\$R\$C\$N\$S\$L\$P\$M\$E\$N\$G\$N\$S\$I\$E\$M\$R\$L\$R\$K\$Q\$I\$K\$G\$-V\$T\$I\$F\$G\$R\$L\$R\$V\$H\$G\$R\$M\$L\$S\$I\$-D\$Y\$D\$	1811
Qy	722	T\$N\$G\$R\$-Q\$O\$T\$D\$P\$K\$V\$H\$L\$S\$M\$S\$K\$Y\$D\$N\$N\$Q\$-----I\$A\$N\$T\$H\$S\$Y\$-----G\$V\$S\$E\$K\$	759
D\$	1812	R\$N\$I\$T\$E\$K\$I\$Y\$D\$H\$R\$F\$T\$L\$R\$I\$Y\$D\$Q\$G\$R\$F\$P\$M\$L\$P\$S\$G\$L\$A\$V\$A\$N\$V\$S\$Y\$F\$N\$G\$R\$L\$A\$Q\$G\$A\$M\$S\$R\$	1871
Qy	760	I\$T\$V\$D\$P\$T\$T\$A\$T\$K\$Q\$L\$Q\$S\$N\$N\$V\$Q\$T\$E\$K\$E\$V\$T\$Y\$T\$P\$S\$Q\$P\$I\$Q\$T\$L\$F\$D\$E\$A\$H\$L\$Q\$S\$C\$H\$T\$L\$T\$R\$D\$G\$M\$	819
D\$	1872	T\$D\$I\$D\$-----K\$Q\$R\$I\$V\$S\$R\$M\$P\$A\$D\$K\$-W\$S\$Y\$S\$Y\$D\$L\$R\$K\$S\$W\$V\$L\$Q\$S\$O\$R\$Q\$-----	1910
Qy	820	R\$V\$R\$E\$T\$A\$I\$Q\$C\$T\$I\$Y\$O\$Y\$D\$N\$N\$R\$V\$I\$Q\$T\$L\$P\$O\$E\$T\$V\$N\$R\$K\$A\$P\$S\$P\$T\$T\$L\$T\$D\$I\$R\$V\$N\$G\$S\$I\$Q\$Q\$	879
D\$	1911	-----I\$F\$E\$Y\$D\$S\$D\$R\$L\$A\$V\$M\$P\$S\$-----V\$A\$R\$H\$-S\$M\$H\$T\$S\$I\$G\$Y\$R\$-----N\$T\$N\$P\$	1948
Qy	880	T\$F\$D\$G\$L\$S\$R\$L\$T\$O\$S\$Q\$D\$G\$R\$W\$A\$Y\$T\$Y\$S\$A\$N\$Q\$C\$P\$S\$T\$V\$T\$P\$D\$G\$F\$I\$H\$Y\$Q\$-Q\$P\$E\$L\$D\$A\$V\$L\$Q\$V\$A\$S\$	937
D\$	1949	P\$E\$S\$N\$A\$V\$I\$P\$D\$Y\$S\$D\$D\$G\$R\$L\$-----K\$T\$S\$F\$L\$G\$R\$Q\$V\$F\$Y\$K\$G\$L\$S\$K\$S\$E\$I\$Y\$V\$D\$-S\$	1994
Qy	938	N\$E\$I\$T\$Q\$S\$S\$Y\$N\$P\$V\$G\$A\$L\$-L\$K\$A\$V\$E\$Q\$S\$L\$T\$P\$Y\$Y\$S\$G\$L\$K\$M\$E\$N\$D\$M\$K\$M\$S\$T\$M\$T\$L\$R\$G\$E\$N\$	995
D\$	1995	T\$A\$V\$I\$-F\$G\$Y\$E\$T\$T\$G\$V\$L\$K\$M\$V\$L\$Q\$S\$G\$F\$S\$C\$T\$T\$R\$Y\$R\$K\$I\$G\$P\$L\$-----V\$K\$Q\$I\$R\$F\$S\$E\$G\$V\$N\$	2045
Qy	996	G\$Y\$T\$L\$T\$Y\$T\$O\$I\$K\$I\$S\$H\$T\$G\$R\$V\$T\$O\$I\$D\$S\$S\$I\$K\$T\$L\$-----N\$Y\$D\$L\$N\$--B\$H\$I\$G\$S\$-----Q\$V\$T\$	1041
D\$	2046	A\$R\$P\$Y\$T\$-----Y\$H\$D\$N\$F\$R\$I\$A\$S\$I\$K\$P\$V\$S\$E\$P\$L\$P\$D\$L\$R\$Y\$B\$E\$I\$S\$K\$V\$H\$F\$G\$F\$G\$Y\$Y\$D\$N\$	2099
Qy	1042	D\$L\$A\$T\$G\$M\$L\$T\$T\$V\$E\$F\$D\$-G\$L\$N\$R\$E\$G\$R\$K\$L\$C\$D\$S\$G\$H\$T\$L\$D\$I\$Q\$S\$M\$L\$K\$T\$Q\$L\$A\$N\$R\$I\$V\$K\$L\$N\$--G\$V\$	1096
D\$	2100	Q\$I\$T\$T\$A\$W\$T\$L\$S\$K\$H\$D\$T\$G\$R\$I\$K\$E\$V\$O\$Y\$E\$M\$F\$R\$S\$I\$M\$W\$T\$V\$Q\$Y\$O\$S\$M\$-----G\$R\$Y\$I\$K\$E\$L\$K\$G\$P\$	2153
Qy	1097	L\$Q\$R\$E\$O\$Y\$S\$Y\$R\$N\$R\$N\$Q\$Y\$K\$C\$D\$G\$A\$-----E\$C\$P\$D\$K\$G\$H\$S\$I\$V\$T\$Q\$N\$F\$Y\$D\$I\$G\$N\$T\$A\$C\$H\$T\$T\$	1149
D\$	2154	Y\$A\$N\$T\$T\$K\$T\$Y\$D\$-----Y\$D\$G\$D\$Q\$Q\$S\$A\$V\$A\$N\$D\$R\$P\$W\$R\$Y\$-----S\$T\$D\$L\$A\$N\$L\$-----	2191
Qy	1150	F\$A\$D\$T\$E\$H\$A\$T\$E\$F\$K\$A\$P\$N\$P\$C\$Q\$L\$T\$E\$V\$H\$H\$T\$P\$D\$M\$P\$N\$I\$K\$Y\$D\$K\$A\$G\$R\$V\$N\$I\$-----T\$D\$N\$G\$	1203
D\$	2192	-----H\$L\$N\$P\$G\$N\$S\$V\$R\$L\$-----M\$P\$-----L\$R\$Y\$D\$L\$R\$R\$I\$R\$T\$R\$L\$D\$G\$V\$O\$Y\$K\$ID\$D\$G\$	2222
Qy	1204	-----N\$T\$E\$N\$F\$Y\$T\$L\$G\$R\$L\$-----Q\$N\$Q\$G\$S\$Y\$Y\$G\$D\$P\$L\$N\$R\$V\$S\$Q\$T\$D\$T\$D\$L\$C\$E\$L\$Y\$R\$E\$T\$M\$V\$	1253
D\$	2228	Y\$L\$C\$O\$R\$G\$S\$D\$I\$E\$Y\$N\$S\$K\$G\$L\$T\$R\$A\$Y\$N\$K\$A\$S\$M\$S\$V\$O\$Y\$R\$D\$G\$G\$R\$A\$S\$Y\$K\$T\$N\$-L\$G\$H\$N\$I\$Q\$F\$Y\$S\$D\$L\$H\$	2286
Qy	1253	N\$E\$V\$R\$-----N\$E\$M\$I\$R\$L\$--R\$T\$G\$E\$T\$I\$A\$O\$Q\$R\$A\$S\$K\$V\$L\$T\$R\$D\$S\$Q\$S\$--V\$I\$T\$S\$D\$K\$Q\$U\$L\$S\$	1300
D\$	2287	N\$P\$R\$I\$T\$H\$V\$N\$H\$S\$N\$E\$I\$S\$T\$Y\$D\$D\$G\$H\$L\$F\$A\$M\$E\$S\$S\$G\$E\$E\$Y\$A\$S\$D\$N\$T\$G\$P\$L\$A\$V\$S\$I\$N\$G\$L\$M\$I\$K\$	2346
Qy	1301	Q\$E\$A\$V\$S\$A\$Y\$K\$-H\$K\$S\$T\$A\$N\$D\$A\$S\$I\$L\$G\$Y\$N\$G\$R\$A\$D\$P\$V\$S\$G\$V\$T\$L\$G\$N\$G\$R\$S\$Y\$D\$P\$L\$M\$R\$E\$H\$T\$P\$D\$--	1354
D\$	2347	Q\$L\$Q\$Y\$T\$A\$G\$E\$I\$Y\$O\$S\$N\$P\$F\$O\$W\$I\$G\$H\$G\$G\$L\$Y\$D\$P\$L\$T\$K\$L\$V\$H\$F\$T\$Q\$--R\$D\$Y\$D\$V\$L\$A\$G\$W\$T\$S\$P\$D\$Y\$T\$M\$	2404
Qy	1355	-----S\$L\$S\$P\$G\$A\$G\$I\$N\$P\$Y\$S\$C\$L\$D\$G\$P\$N\$R\$S\$-D\$P\$S\$H\$L\$-----S\$M\$Q\$M\$W\$T\$I\$-----	1392
D\$	2405	K\$W\$N\$G\$E\$P\$A\$P\$F\$-----N\$L\$Y\$M\$E\$K\$S\$N\$N\$P\$S\$E\$L\$D\$L\$K\$N\$Y\$T\$D\$V\$K\$S\$W\$L\$V\$M\$F\$G\$O\$L\$S\$N\$I\$T\$P\$G\$P\$P\$R\$A\$	2455
Qy	1393	-----Q\$W\$G\$I\$A\$G\$L\$L\$T\$T\$A\$T\$G\$M\$A\$T\$A\$A\$G\$-----I\$A\$A\$I\$A\$-----	1421
D\$	2460	K\$M\$Y\$V\$P\$P\$E\$V\$L\$S\$Q\$A\$S\$E\$N\$Q\$L\$T\$G\$V\$Q\$T\$T\$E\$R\$H\$Q\$A\$F\$M\$A\$L\$E\$G\$V\$Y\$T\$K\$K\$A\$S\$I\$E\$K\$A\$G\$H\$W\$	2519
Qy	1422	S\$T\$T\$T\$M\$A\$G\$A\$L\$-----S\$Y\$S\$T\$S\$T\$V\$S\$G\$A\$L\$E\$D\$A\$P\$K\$A\$S\$S\$I\$T\$G\$-----	1459
D\$	2520	F\$A\$T\$T\$P\$T\$I\$G\$K\$M\$A\$E\$K\$G\$R\$V\$T\$G\$V\$S\$S\$I\$A\$S\$--E\$D\$-S\$R\$K\$A\$S\$V\$A\$N\$N\$A\$Y\$Y\$D\$K\$H\$Y\$S\$I\$E\$G\$	2575

QY 1460 -----WVSMGMAAGLA-----ESA-----IKGKTALATLGAFAEDGE 1493  
DB 2576 DTHYFVIGLGSADGDLVLTGTTIGRKVLBSGVNTVVSQPTLLVNGKTRFRFINI-----EFQY 2631  
QY 1494 NALIKSTSESRIRIKGVT-RSLDREIVRMEGOVIKDHSGYTDNFMKGEOAILVHGDK 1552  
DB 2632 STILLS-----IRYGLVPTLDEKAR-----VLQARQALGTMAKEQQAADG-R 2678  
QY 1553 DGFVHTGK 1563  
DB 2679 EGSRLWTEGK 2689  
RESULT 14  
US-10-453-372-154  
; Sequence 154, Application US/10453372  
; Publication No. US2006003323A1  
; GENERAL INFORMATION:  
; APPLICANT: Alabrook, et al.  
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD  
; FILE REFERENCE: 21402-589 A  
; CURRENT APPLICATION NUMBER: US/10/453,372  
; PRIOR APPLICATION NUMBER: 2003-06-03  
; PRIOR FILING DATE: 2001-02-23  
; PRIOR APPLICATION NUMBER: 60/185967  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: 09/823187  
; PRIOR FILING DATE: 2001-03-29  
; PRIOR APPLICATION NUMBER: 60/195792  
; PRIOR FILING DATE: 2000-03-10  
; PRIOR APPLICATION NUMBER: 09/839446  
; PRIOR FILING DATE: 2001-03-19  
; PRIOR APPLICATION NUMBER: 60/199476  
; PRIOR FILING DATE: 2000-03-25  
; PRIOR APPLICATION NUMBER: 09/863776  
; PRIOR FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: 60/208263  
; PRIOR FILING DATE: 2000-05-31  
; PRIOR APPLICATION NUMBER: 09/939398  
; PRIOR FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: 60/227800  
; Remaining Prior Application data removed - See file Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 1609  
; SOFTWARE: Cnaseq1 version 0.1  
; SEQ ID NO 154  
; LENGTH: 2733  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-453-372-154  
Query Match 3.1%; Score 272.5; DB 6; Length 2733;  
Best Local Similarity 19.5%; Pred. No. 3e-09;  
Matches 365; Conservative 251; Mismatches 652; Indels 603; Gaps 99;  
QY 49 GHIYNG-NLBPPLPLTSLYSPLNTDIFGFGFGLSVYDRKKSLLSTGENYK--- 104  
DB 1066 GHLFOKSFQASPNLSTFIW---DKTDA-----YGGRVYGLSAVVSV--GPEYETCP 1113  
QY 105 ---VIETDKTVKLOQKLD--NL---RPEK---DLKENCYRIHK-SGDIEVLTG----- 147  
DB 1114 SLIMEKRTALLQGEELDPNSLNGWLDKHLNLYKSG---ILHGCTENGFLTQOPAI 1170  
QY 148 ---FNNAFDLKV-----KKLNPAGHAIYIMNFATOPRLNRIYDDLDGDHDP 195  
DB 1171 TSIIMGRRRISICSPSCNGLAEGNKLLAVALAVGIDGSL----- 1210  
QY 196 LLNLEYGQGLIKTLLLPFGQKGYRTELRFNLROL-----NSIHNSLGNENPLTWSFGYT 251  
DB 1211 ---YVDENYIRIRIFPSRNVTSILELR--NKEFKHSNNPAHKYLLA--VDEVSGLLYVS 1262  
QY 252 PIKNGILGOWITTSWTAPGGLKETVNVYNNNGHHFPOSANLPLVLRVYTLMLKQVAGACGP 311

DB 1263 DTNSRRILY-----RVSLSGTKOLAGSEVAVG-----TGEQCLPF---DEARCGGK 1308  
QY 312 AIDAEISYSHNVYGGGNGINW-----NKDNLGLMTENVYSGTESRRYK---DK 360  
DB 1309 AIDA--TLMSPRGIAVDKMLMFEVDTAMIRKQD--NGIISTL--LGSNDLTLAVAPLSCDS 1364  
QY 361 EGHDIYRIE-----RTYNNVHLITSECKOONGYIOTETAPYVAIIGHNFPDSQSF 412  
DB 1365 SMDVAQVLEWPTDLAVNPNDSLYVL-----ENNVLRITENHOVSIIA---GRPMHC 1415  
QY 413 QLPKTKETWRSADNSYRSBITEETPDESGNPLTKYIKDKKTOKIISPSTHMEYYPAGE 472  
DB 1416 QVPGIDVSLKLAHSLMESASAIASHTGVLYTETDEKKINRLQVTNNGEICLLAGA 1475  
QY 473 VDNCPPE-----PYGTRPVKKIIGRPYDEFFDDP 503  
DB 1476 ASDDCNDVNCNCYSGDDAYATDAILNPSLSLAVAPDG--TIYIADGNIRIRAVSKXP 1534  
QY 504 --EKFIQYRSLIGSGHVTLKIEERHYSATQLNSTL---FOYNTDKSELRLKQTEC 558  
DB 1535 VLNAPNOYEAASRGEQLYFPNADGIHQYVSLVTSGLVNFYSTD--NDVTELI----- 1588  
QY 559 TKGNGKTVSVHKKFTYTKODTLQOSH-----SITTHDNFTIHSQ----- 600  
DB 1589 --DNNGNSLKI-----RDSGMPRHLLMPDNOIITLTVGTNGGLKAVSTQMLBGLM 1639  
QY 601 VRSRYTGLRPSDDTDKDIYQMSYDLGRLTRLTNGST-----PANTLTVOYELNLT 654  
DB 1640 TYDQNTGLATKSGSETGWTFFYDHDGKRLTNVTRPGVVTSLHREMEKSTIIDIEENS 1699  
QY 655 QDDNRPFVITTPDVNG-----NQLRNEFD-----GAGRH----- 684  
DB 1700 DDD-----VYITNLSVVASLYTVQDVANSYGLCANNGLTRVNVYANGMGSFSEPHVL 1754  
QY 685 -----VSQC-----LKSDGDKFTYTHTQOYDQGRHHTSYSDYL 721  
DB 1755 AGTTPTRGNCINSLPWENGINSIEMRLRKEQIKG--VTIFGRRLVHGRMLSI--DYD 1811  
QY 722 TNGR-QQTDPDKHLSSKSYDNMGQ-----IANTMSY-----GVSEK 759  
DB 1812 RNIRTEKIYDHRKFTIRIYDOGRPPLMPSSGLAIVANVSYPFNGRLAGLGANSEK 1871  
QY 760 ITVPITLTATKLOQSNNNVQTSKEVTYTPSQOPIQITLFDAGHLQSGHTLTRDGM 819  
DB 1872 TDID-----KQRIYSKMFADCK--VMSYSYLDKSVLLQSGROY----- 1910  
QY 820 KVRKETDAIQCTIYQYDYNRNVIQITLPDGTIVNRKYAPSTDTLITDIFNGISLQOQ 879  
DB 1911 -----IFEYDSSDRLLAVTMP--VARH--SMSTHTSIGYIR-----NIYNP 1948  
QY 880 TFDGLSRLTQSGDGRWATYYSAGNDQCESTVITPPGQFIHQY--QPELDVAVLQVAS 937  
DB 1949 PESNASYIFEDYSDDGRIL-----KTSPLGTGRQVYFGYKLSKSEIYYD--S 1994  
QY 938 NEITQOSYNPVTGAL--LKAVALGOSLTPYPSGLKMEINDMKMSYLAWLKLEN 995  
DB 1995 TAVT--FGYDTEVLKVMNLQSGGFCITIRKIGPL-----YDQIIRFSEGVN 2045  
QY 996 GYTDLTGTQIKISDTHGRVTOIQDSSIKTTL-----NYDDL--RHIGS-----OYT 1041  
DB 2046 ARPDYT-----YHDSFRIASIKRVISSEPLPDLYKAYDEISGVEHFGFGLVYDIN 2099  
QY 1042 DLATGHMLTTVERD--GLAREIKRCLDSGHTLLDIOQSLKTOQLANRLVTKL--GV 1096  
DB 2100 QITTAWTLISKHPTDGRIRIKVEYEMFRSLMYMTVOYDSM-----GRVYKELKGP 2153  
QY 1097 LQRTQOYSYDSRNLNQYKCDGA-----ECPDTKYGSHSVTQNFYDIYGNITACHYT 1149  
DB 2154 YANTTKITYD-----YDGGQLOSAVANDRPTWRY-----SYDIANGNL----- 2191  
QY 1150 FADGTEDHATPEKANPTDPCQLTEVNHHTHPMDPNIRIKYDKAGRVINI-----TDHG 1203

```

Db 2192 -----HLINPNSVRL-----MP-----LRYLDRIRTRLGVQKIDDDG 2227
Qy 1204 -----NTNFTYDTLRL-----QNGQSVYGVDPPLNRLVSOQTDPLDDELYRRTMLV 1252
Db 2228 YLCORSGDITFYNKSKGLTRAYNKASGWSVOYRDGGRASXYKTN-LGHHLQYFYSDLH 2286
Qy 1253 NEVR-----NGEMIRLL--RTGETIIAQORASKVLTLGTSTQOS--VILTSDKNTLS 1300
Db 2287 NPIRITHVNHNSSEIRSLVYDLOGHLFAMESSSGEYVYASDNTGTPPLVPSINGLMIX 2346
Qy 1301 QEAYSAYGK--HKSTANDASILGNERADPVSGVTHLNGYNSYDPTLMRFTPTD--- 1354
Db 2347 QLOVTAAGEIYYDSNPFQWVIGFHGGLYDPLTGLVHFTQ--RDYDLAAGRTMSPDYTM 2404
Qy 1355 -----SLSPFGAGINPYSCLDPIKRS--DPSGLH-----SQWAMTGI----- 1392
Db 2405 KMWGKEBAPF-----NLMPKSNPNLSSELDLKNYIVDVKSWLMVFGPQLSNII PGPPRA 2459
Qy 1393 -----GMGIAGLLLTATGMAIAAG-----IAATA----- 1421
Db 2460 KMYFVPPPYELSESQASENQOLITGVQOTTERHQAEMALEGVYITKGLASIREKAGW 2519
Qy 1422 -STSTTALAFAL-----SVTSDITSIVSGALEDASPKASITG----- 1459
Db 2520 FATTPTPIGKIMPAIKEGRVTTGVSSIAS--ED-SRKVASVLMNAVYLDKMHYSIEGK 2575
Qy 1460 -----WYSMGGAAGLA-----ESA-----IKGTGLATHLAPADGE 1493
Db 2576 DTHYFVAGISADGLVTLGTTIGRKVLSEGVNVTVSQPTLLVNGRTFRFTNI--BFQY 2631
Qy 1494 NALLKSTSESSRIKMGVT-RSLDEIYRNEEGVYIKHSGRYDNFMGKEQOAILVYGDK 1552
Db 2632 STILLS-----IRYGLTPTLDEKAR-----VLDQARALGTMAKEQOKARG-R 2678
Qy 1553 DGFLYHTEGK 1563
Db 2679 EGSRMLTEGK 2689

```

RESULT 15  
US-10-453-372-116  
Sequence 116, Application US/10453372  
Publication No. US2006000323A1  
GENERAL INFORMATION:  
APPLICANT: Alsobrook, et al.  
FILE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD  
FILE REFERENCE: 21402-589 A  
CURRENT APPLICATION NUMBER: US/10/453,372  
CURRENT FILING DATE: 2003-06-03  
PRIOR APPLICATION NUMBER: 09/789390  
PRIOR FILING DATE: 2001-02-23  
PRIOR APPLICATION NUMBER: 60/185967  
PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: 09/823187  
PRIOR FILING DATE: 2001-03-29  
PRIOR APPLICATION NUMBER: 60/195792  
PRIOR FILING DATE: 2000-03-10  
PRIOR APPLICATION NUMBER: 09/839446  
PRIOR FILING DATE: 2001-03-19  
PRIOR APPLICATION NUMBER: 60/199476  
PRIOR FILING DATE: 2000-03-25  
PRIOR APPLICATION NUMBER: 09/863776  
PRIOR FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: 60/208263  
PRIOR FILING DATE: 2000-05-31  
PRIOR APPLICATION NUMBER: 09/939398  
PRIOR FILING DATE: 2001-08-24  
PRIOR APPLICATION NUMBER: 60/227800  
PRIOR FILING DATE: 2000-08-25  
Remainder prior application data removed - See file wrapper or PALM.  
NUMBER OF SEQ ID NOS: 1609  
SOFTWARE: Cureseqlist version 0.1  
SEQ ID NO 116

```

; LENGTH: 2765
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-453-372-116
Query Match
Best Local Similarity 19.5%; Pred. No. 3e-09;
Matches 365; Conservative 251; Mismatches 652; Indels 603; Gaps 99;
3.1%; Score 272.5; DB 6; Length 2765;

```

49 GHIVANG-NIGPLPLPLTSLPLNKTDIGFICFNGLSYDRKNSLSLSTGBNTK--- 104  
1098 GHIFOKSPQASPMIASTFIW---DKTDA-----YGORVYGLSDAVSV--GFREYETCP 1145  
105 ---VLETDKYKLOQKLD--NL--RFEK-----DLKENCRIIHK--SGDIEVLG----- 147  
1146 SLTLMERKTLALOGFELDPNSLGGMSLDKHLILNVASG--LHKFGENQFLTQOPAI 1202  
148 -----FNNAFPLKVP-----KKLNPAGHAIYIDWNEFATQPLNRIYDDLGDHDP 195  
1203 TSIMNGRRRSISICPGCNGLABGNKLLAPVALAVGIDSL----- 1242  
196 LNLBYOGLIKITLTPPGQKSGYRTBLRPLNQL-----NSINFSLGNEPLTWSEFGYT 251  
1243 -----YGDGFNYIRRIFFPSRNVTSILELR--NKEFKSNPAPAKHYLLA--VDPVSGSLYVS 1294  
252 PIGKNGILQWITSMTAPGLKETVNVSNNOGHHPQSANLVLPLYTLMKQVPGAGP 311  
1295 DINSRRIRY-----RVKSLSGTKDLAAGSEVAG-----TGEOLCP--DEARGDGK 1340  
312 AIQAEYSYSHNVYGGSGNGINW-----NKLDNLVGLMTEVNYGSTESEARYK---DK 360  
1341 AIDA--TMSPRGIAVDKNGMLFVDATMIRKQD--NGIISTL--LSGNDLTARPLSCDS 1396  
361 EGHDIQVRIE-----RTYNNYHLTSECKQONGYIQTETAYAIIGHNFDOSQSF 412  
1397 SMDVAGVRLWEPFDLAVNPMDNLSLYL-----ENNVILIRITENHQVSIIA--GRPMHC 1447  
413 QLPKTEETWRGADNSYRSIEITETDESGNPLTKYIKKKTKIISPTHMEYPPAGE 472  
1448 QVPGIDYSLSKLAHLSBASASALASHTGVLVITETDEKIRLRLQVTTNGEICLLAGA 1507  
473 VDNCPPE-----PYGFRFVKKIQTYPYSEFPDOP 503  
1508 ASDCDCKNDVNCYSGDADAYATDALINSSSLAVAPDG--TIYIADLGNIRIARVSNKR 1566  
504 --EKFIQYRYSILGSOHVTLKIEBHYSATOLLNSTL---FOYNTDKSELGLKQTEC 558  
1567 VLNAFQYEAASPGQBELVFNADGHHQYVSLVGTGEVLNFTYSTD--NDVTELI----- 1620  
559 TKENGKITSVNHKFTYTKODDTLQOSH-----SITHDNFTIHSQ----- 600  
1621 --DNNGNSLKI-----RRDSGMPRHLLMPDNOIITLVGNGGLKVVSTNLEBGLM 1671  
601 VRSRYTGRLEPSTDOTDIYQWMSYDKLGRLLTRTLNSGT-----PYANTLVYDVELNLT 654  
1672 TYGNTGLATKSDDEGWTTFYDIDHGRLLTNVTRPTGVVTSIAREMEKSITIDIENSNR 1731  
655 QDDNRPFFVITTTDVG-----NOLNNEPD-----GAGRH----- 684  
1732 DDD-----VTVINLSVSEASVTVQDVQVNSYQLCNGNGLRVMYVANGMISHSBPHVL 1786  
685 -----VSQC-----LQSDSDGFGYTIHTQOYDEQGHHTSYSDYL 721  
1787 AGTITPTIGRCNISLPMENGANSIEWRLRKQIKG--VIFGKRLRVHGNLSLI--DYD 1843  
722 TNGR--OOTPDKTHLSMSKSYDNWGO-----LANTHWSY-----GVSEK 759  
1844 RNIIRTEKIYDHRKFTLRIRIYYQVGRPLMLBSSGLAAVNVSTFPNGRLAGLORGAMSER 1903  
760 ITVDPTTLATKQLQGSNNNVOTGKEVTTYTPSQPQIQLTFDEAGHLSCHTLTBDGWD 819  
1904 TDID-----KGRIVSRMFADGK--VMSYSYDKMSMVLILQSQRY----- 1942

```

QY 820 RVRKETDAIGQCTIYQYDYNVRVIOITLPDGTIVNRKAPFSTDTLLITDIRVNGISIGOO 879
Db 1943 -----IPEYDSSRLLAVTMPB---VAHH--SMSTHTSIGYR---NINP 1980
QY 880 TFDGLSRLTOSQDGRVWAAVYYSAGNDCEPSTVITPDGQFIHYQ--OPELIDAVLOVAS 937
Db 1981 PESNASVIFPDSDDGRIL-----KTSFLGTGRQVFYKYGKLSKLSSEIYVD--S 2026
QY 938 NEIQOQPSYNPVNGAL--LKAVAEGQSLPIYPSGRLXNENINDMKKMSYLWTLRGLN 995
Db 2027 TAVT--FGYDETTGLVKKNVIOGSGFSCCTIRKIGPL-----VDQITRSEEGMN 2077
QY 996 GYTDLTGTIOKISHDTHGRVTOIKDSSIKTYL-----NYDDL--RHIGS-----QVT 1041
Db 2078 ARFPDY-----YHDSFRISIKFVISEPLPVDIVRYDEISGKVEHFGKEGVYYIDIN 2131
QY 1042 DLATGHMLTTTVEPD--GLNREIRKLCDSGHTLIDIQOSWLKTOQLANRIVKLN---GV 1096
Db 2132 QIITTAVWTLSKHFDTHGRIKGVQYEMFRSIMMMTWQYDSM-----GRVIRELKGIP 2185
QY 1097 LQRTQOYSYDSRNLNQQKCGA-----ECPTDKYGHSLVTQNFYDIIGNITACHTT 1149
Db 2186 YANTTKYTYD-----YDGGQLOSAVANRPTRRY-----SYDLGNL----- 2223
QY 1150 FADGTEBHATPKFANPTDPCQLTEVHTHPMPDNIRLKYDKAGRVIINI-----TDNHG 1203
Db 2224 -----HLNPGNSVNL-----MP-----LRVLDRLRITRLGDVQYKIDDDG 2259
QY 1204 -----NTENPTYDTLGRIL-----ONGQGSVYGYDPLNRLVSQKTDPLDCELYRETMLV 1252
Db 2260 YLCORGSDFEYNKSGKGLTRAYNKASGWSVQRYDGVRRASYKTN--LGHNLQYFYSDLH 2318
QY 1253 NEVR-----NEMIRLL--RTGETIIAQORASKVLLTGTDSQGS--VILTSDKONLS 1300
Db 2319 NPTRIITHVYNSNSBITSLYYDLOCHLFAMESSGEERYVASDNGTPLAVFSINGLMIX 2378
QY 1301 QEAYSAYGK--HKSTANDASIIQYNGERADPVSGVTHLNGYRSYDPTLMRFHTPD--- 1354
Db 2379 QLOYTAVGEIYYDSNPDQWVIGFPHGLYDPLTKLVHFTQ--ROYDVLAGRMTSPDYTMW 2436
QY 1355 -----SLSPFGAGINPYSYCLGDPINRS-DPSGHL-----SWQAWTGI----- 1392
Db 2437 KNVKEPAPF-----NLTFMFKSNPFLSELDLKNYVTDVKSWLVWFGFOLSNIIIPGPRA 2491
QY 1393 -----GMGIAGLLLTIAIGMAIAAGS-----IAAIA----- 1421
Db 2492 KMIVVPPPEYELSESQASENGQILITGVQOTTERHNOAFWALEGOVITTKLHASIREKAGW 2551
QY 1422 -STSTTALAFGAL-----SVTSDITSIVSGALBEDASPKASSILG----- 1459
Db 2552 FATTTPIIGKIMFAIKGRVYTGVSJAS--ED-SRKVASVLNNAYYLDKMHYSIEGK 2607
QY 1460 ----WVSKMGGAAGLA-----ESA-----IKGCTKLATH/GAFAEDGE 1493
Db 2608 DTHYFVKIGSADGLVLTGTTIGRKVLSEGVNVTVSQPTLLVNGTRRFTNI---EFOY 2663
QY 1494 NALKSTSESRIKMGVY--RSIDREIVNNEGQVTKHDSRGYTDNFMKGEOALIVHGD 1552
Db 2664 STLIS-----IRYGLTPDTLDEKAR-----VLQARQALGTAMAKEOQKARDG-R 2710
QY 1553 DGFVYHTEGNK 1563
Db 2711 EGSRLMTEGK 2721

```

Search completed: January 30, 2006, 09:53:34  
 Job time : 29.3373 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2006 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: January 30, 2006, 09:41:43 ; Search time 29.1754 Seconds

(without alignments)  
5517.344 Million cell updates/sec

Title: US-09-889-874A-23

Sequence: 1 VYIKFKLRRRTMSDNER.....PKIIIGRTKVKPKTPRP 1673

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 263416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 263416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR 80: \*  
2: PIR: \*  
3: PIR: \*  
4: PIR: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1214	13.7	709	2 S38241	hypothetical prote
2	912.5	10.3	528	2 S38242	hypothetical prote
3	432	4.9	2334	2 S32920	cell wall-associat
4	429.5	4.8	336	2 S38239	hypothetical prote
5	385	4.3	2167	2 AP1489	cell wall-associat
6	381	4.3	2183	2 T37218	hypothetical prote
7	365	4.1	1400	2 E90886	RhsB core protein
8	353.5	4.0	843	2 AB0539	Rhs-family protein
9	342	3.9	1397	2 C64805	RhsB core protein
10	341.5	3.8	1404	2 E90658	RhsB core protein
11	341	3.8	1394	2 H91236	RhsB core protein
12	340.5	3.8	1411	2 E65155	RhsB core protein
13	339	3.8	1426	2 H64780	RhsB core protein
14	336.5	3.8	1404	2 E85509	hypothetical prote
15	331.5	3.7	1377	2 C65159	RhsA protein precu
16	331	3.7	1409	2 F91187	RhsA core protein
17	326	3.7	1399	2 A99720	RhsC core protein
18	322.5	3.6	1377	2 E86034	RhsA protein in rh
19	319	3.6	1397	2 A85570	RhsC protein in rh
20	316	3.6	1398	2 E85549	hypothetical prote
21	313	3.5	1398	2 H90698	RhsD core protein
22	297	3.3	1512	2 AH0439	probable membrane
23	280	3.2	1317	2 E83310	conserved hypochet
24	278	3.1	985	2 E86084	hypothetical prote
25	276.5	3.1	1438	2 A10093	conserved hypochet
26	271.5	3.0	1354	2 AG0538	Rhs-family protein
27	266.5	3.0	2515	2 S47008	tenascin-like prote
28	248.5	2.8	1616	2 E90704	Rhs core protein w
29	245.5	2.8	1645	2 H85554	hypothetical prote

30	245	2.8	2825	2 T14271	DccA protein, stre
31	230	2.6	794	2 T36972	probable membrane
32	230	2.6	849	2 T46253	hypothetical prote
33	220.5	2.5	1185	2 A42404	collagen adhesin -
34	219.5	2.5	1962	2 A32634	lactocerin (EC 3.4
35	218.5	2.5	2406	2 A54148	odx protein - fru
36	217	2.4	356	2 T37136	hypothetical prote
37	216.5	2.4	2894	2 C64474	hypothetical prote
38	215.5	2.4	2044	2 AB1180	probable peptidogl
39	213	2.4	1487	2 AG2560	hypothetical prote
40	211	2.4	3283	2 AC1018	large repetitive p
41	208	2.3	2703	2 H81193	hemagglutinin/hemo
42	207	2.3	2893	2 A64556	toxigen-like outer m
43	205	2.3	1902	2 S05997	lactocerin (EC 3.4
44	203.5	2.3	1366	2 S57664	IGA-specific metal
45	202.5	2.3	1349	2 A11476	cell surface prote

## ALIGNMENTS

## RESULT 1

S38241  
hypothetical protein - Coccidia burnetii

C:Species: Coccidia burnetii

C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 09-Jul-2004

C:Accession: S38241

R:Thiele, D.; Willem, H.; Haas, M.; Kraus, H.

submitted to the EMBL Data Library, October 1993

A:Reference number: S38215

A:Accession: S38241

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-709 &lt;THI&gt;

A:Cross-references: UNIPROT:Q45948; UNIPARC:UP100000BA5E9; EMBL:X75356; NID:G407370; PIR:

Query Match	13.7%	Score 1214;	DB 2;	Length 709;
Best Local Similarity	40.8%	Pred. No. 2e-59;		
Matches 287;	Conservative 101;	Mismatches 280;	Indels 36;	Gaps 16;
QY	554	KQTECTKGENKQTYSVVHKFTYTKODDTLQOSHITTHDNFTIHSQVRSRYTGRLESDT	613	
DB	3	RRAEVLTSKKEGKKYQNGTTFPAISSQAHHLQKIDFTGDDGKITSISEQGRYSGLLST	62	
QY	614	DTKDIYQMSYDKLGRLLRTLN-SCTPYANTLTLYELANLQDDNRPPVITTTDVNGN	672	
DB	63	DEIGNVYQYDEIGRLITQVNASSTYASTYRYSLETDARGKYATAKTYTDPKGN	122	
QY	673	QLNEPFGAGRHYSQCLKSDG----DGKFTYHTQYDEQGRHHTSYSDYLTNGRQ	727	
DB	123	QLKTYDGLGRNLKQERLDDAAVSOQKTGWYITTHQOQDALGRSKITTDVL-----R	177	
QY	728	TDDPKVAH---LSMSK--SYDNMGQIANTHMSYGVSEKIVDPITLTATKQLOSNMNVQ	781	
DB	178	LDBEVGHAGSVLSTSKYMHDSWGQNLTVFSQGYQSRSYDDPITRAT--LDPSGSGQK	235	
QY	782	TGKAVTTTPSPQPIQTLTFLDEAGHLQSGHTLRGDRVRSKTDAGCCTIYQYDNYNR	841	
DB	236	LGQQLTEYNLAGPIKTYQYDSQTEGSAHYEDGIGQURKEDLGQITLYEYDHFGR	295	
QY	842	VIGITLPDGTIVRKAPFSTDLTDIRVNGISLGQGTGDSRLTQSDGGVMAVYTY	901	
DB	296	VIGITLPDGTIVRKAPFSTDLTDIRVNGISLGQGTGDSRLTQSDGGVMAVYTY	901	
QY	902	SAGNDQCPSTVITPDGQFIHYQYQPELDDAVLQVANSNEITQPSYNTGALLKAV-AEG	960	
DB	356	ENMS-SVPAAVTAPVGTGVEYVLEKELGNVAKISABEIIQTMVDYDALTGAMTSATQAG	414	
QY	961	QSLTPITYPSGRLKME-NIND---MKMSYLMTLRGLENQYTDLTGTIQLSRDTHRVY	1016	
DB	415	MIRQMTYPSGLKNETSMPDGAQKSTAYTSLAGAPQSTYVFGVQYQYDEHGRRI	474	
QY	1017	QINDSSIKTLTANLDDLRHIGSQVTDLATGMLTTTYEPFGLRREIGRKLCDSGHTLDI	1076	

Db 475 GIBNDIVSLDYDAFGRFTKQATDKKTGAVLSTTLTYDDLNREIKREISASQSVLVI 534  
Qy 1077 QQSLLKTOQLANRYKLVNGVLTQRTQOYSDSRNLNQKCGACFPDVKHSHVTONFT 1136  
Db 535 EGYTORHNLKERITQRTTLRKEMFAYDSRNRRIETCYNGEARPODPYKALHROFSS 594  
Qy 1137 YDIYGNITACHTTADGTEHDATPEFANPTPCOLTVEYHTHPMPDPIRLKYDKAGRVI 1196  
Db 595 YDALGNMTKOTDSSGG-RNRATYIY-SADIPYQLKNNHSDPKREITLEYDKAGRMI 652  
Qy 1197 NITDNHNTENFTYDTLGRLO-NG--QGSVGYDPLNRLVSQ 1235  
Db 653 R-DEAGRT-LRYDALGRLOQVNGAGAKGGQAYADVLTNLVSQ 692

## RESULT 2

S38242  
hypothetical protein - Coxsiella burnetii  
C:Species: Coxsiella burnetii  
C>Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 09-Jul-2004  
C:Accession: S38242  
R:Heide, D.; Willems, H.; Haas, M.; Krauss, H.  
submitted to the EMBL Data Library, October 1993  
A:Accession: S38242  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-528 <THI>  
A:Cross-references: UNIPROT:Q45949; UNIPARC:UP10000060C0B; EMBL:X75356; NID:g407370; PII  
C:Superfamily: Coxsiella burnetii hypothetical protein

Query Match 10.3%; Score 912.5; DB 2; Length 528;  
Best Local Similarity 40.8%; Pred. No. 6.6e-43;  
Matches 222; Conservative 72; Mismatches 203; Indels 47; Gaps 17;  
Qy 18 NEF-FTQANFTSAVSGGVDPRTGSLYNIQTILGHVNGNLGPTLPLTSLSPINKTDI- 75  
Db 2 NELPYQTQNTFISAVQGVDPRTGSLFTYNNVLAELTGNIDNLGPDLPLFLTANSHLSTNIC 61  
Qy 76 GFGIGFNGSLVDRKSNLSLSSTGENYKVLFTDKTVLQOQKLDNLFEDKENCYRI 135  
Db 62 GFGIGCVSGLISYDKNNLILSSGERYKTDMDSDVYVQKKNPFERK-1KNGYII 119  
Qy 136 IHKSGDIEVLTG--FNNAFPLKVPKCLNPAHAIYIDMVPFATQPLNRYDLDGDH 193  
Db 120 KYKNGKTYLVNKKGDYLF--LPQKLPSPUGMPLKLTWNRQYVNLTKIEDAKD--- 173  
Qy 194 IPLNLLEYQGLIKTILLFPQOKEGYRTLRFPLNRQNLNSHNSLGNENPLTWSPGYTPI 253  
Db 174 -VLCIKIDYQPSDMARITWPGKTSYQLDFVNEIYLVWVTKNSRSR--LWSPNYVDV 230  
Qy 254 GKNGILGMITSMAPGSLKKTAVNYSNNQGHHPQSANLVLPVYTLMKQVPGAGQAI 313  
Db 231 GAGNFT--LQVKSPTGLTETVYQAGV--RPFDESGKALPSVYNYRQSPGQOPDI 285  
Qy 314 QAESYTSHTNYVGGSN--GIWNKLDNLVYG-LMTENYVSGTESRRYDKGCHDIYVIE 370  
Db 286 VKHEIYTSNLTGICASLGKAMNEDENIYVMDIYYSSTEKLIYVNR---LVISIS 341  
Qy 371 RTYNNYHLLTSECKQONGYIOTTETAYVAIIGHNFDSPQSPQLPKTETETWR-SADNSY 429  
Db 342 RIYNSYVLLISETTRQNSCEVAVETDYAKPGLSPDKQPKQFQPKKEKTKWRENSKQC 401  
Qy 430 RSEITETTFDESQNLTVIKDKTKQKILSPSTHMEYTPPAGVDN---CPPEYGGTR 485  
Db 402 RSEITTTTFDEGNLTKIEPD-----GKTEIYVDSKGETDKGIVLPPEPENGVR 454  
Qy 486 FVKKIIQTPYSEF-----KDDPEKIQYRSLIGSQSHVLTKIBERHYSATQLNSTLFG 541  
Db 455 FVKIQTIVPADSEFAPVQQTITAYAOIPICTAGSSLSIAVLQTOETLCSDDL---LIT 510  
Qy 542 YNTD 545

Db 511 INTD 514

## RESULT 3

S32920  
cell wall-associated protein precursor wapa [similarity] - Bacillus subtilis  
C:Species: Bacillus subtilis  
C>Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 09-Jul-2004  
C:Accession: S32920; E69730; T47101  
R:Forster, S.J.  
Mol. Microbiol. 8, 299-310, 1993  
A:Title: Molecular analysis of three major wall-associated proteins of Bacillus subtilis  
protein.  
A:Reference number: S32919; MUID:93302506; PMID:8316082

A:Accession: S32920  
A:Molecule type: DNA  
A:Residues: 1-2334 <EOS>  
A:Cross-references: UNIPROT:Q07833; UNIPARC:UP10000060C90; GB:L05634; NID:g304177; PIDN:CAE  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertek  
C.; Bron, S.; Brouillet, S.; Brunsch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi  
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fultz, C.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Gallier  
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holappel, S.; Hobono, S.; Hullo, M.F.;  
Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel,  
Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle,  
Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,  
A:Authors: Schlecht, S.; Schroeter, R.; Scottone, F.; Sekiguchi, J.; Sekowaka, A.; Serot,  
akeuchi, M.; Yamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Toso, V.; Uchiyama,  
T.; Wintere, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.  
A:Authors: Yoshikawa, H.F.; Zunsstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
A:Reference number: A69580; MUID:98044033; PMID:9384377  
A:Accession: E69730  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-2334 <KUN>  
A:Cross-references: UNIPARC:UP10000060C90; GB:299124; GB:AL009126; NID:g2636442; PIDN:CAE  
A:Experimental source: strain 168  
R:Yoshida, K.; Sano, H.; Seki, S.; Oda, M.; Fujimura, M.; Fujita, Y.  
Microbiology 141, 337-343, 1995  
A:Title: Cloning and sequencing of a 29 kb region of the Bacillus subtilis genome conta  
A:Reference number: 224350; MUID:95219088; PMID:7704263  
A:Accession: T47101  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2334 <YOS>  
A:Cross-references: UNIPARC:UP10000060C90; EMBL:D31856; NID:g603765; PIDN:BAA06656.1; PI  
A:Experimental source: strain BGSICAL  
C:Genetics:  
A:Gene: wapa; N176  
C:Superfamily: cell wall-associated protein wapa  
Query Match 4.9%; Score 432; DB 2; Length 2334;  
Best Local Similarity 20.4%; Pred. No. 4.1e-15;  
Matches 358; Conservative 226; Mismatches 613; Indels 560; Gaps 80;  
Qy 44 IQTLGHIVG-NGNL-----GPTLPLTSLSPINKTDIGFGIGNGLSVYDR 90  
Db 896 IDIPSGQLNGATGVINVEEDLSIDGRPGIGLSRTYNSLSDSDHLTGQGW-----YADA 950  
Qy 91 KNSLISSTGENYVITETDKTVLQOQKLDN-----LRFEDKENCYRIHKSQDIEVL 145  
Db 951 ETSVISTDGGAMY--IDEDATHTRTKADGTYQPPGVYIELTETADQFLTKTD----- 1004  
Qy 146 TGFNNAFPLKVPKCLL-----NPAGAIYIDWN-----FEATQPLNRIYDLDG-- 192  
Db 1005 ----QTNAYFKKGGKGLQKVVDDGNHNAIVTYINDGNQLTATDASGRKLTFTYDE-NGHVT 1060  
Qy 193 -----DIPLNLEY--QGLI-----KTILTVLPG 214

```

Db 1061 SITGPKKAVTSYENDLLKKVTDIDGTATSYDSEGRVLKQVANSSTAEKPFTEY-- 1118
Qy 215 QKGGYRFLAFLRQLNSINFS-----LGNENPLTMSFGTTPIGKNIILQOMITSM 267
Db 1119 QSGHRLERKAINAKKETYYVSYADAKTLLMTOPNGRKVOGYNEANP-----IQVID 1172
Qy 268 APGLAKETVA--YNNN-----QGHFPOSANLPVLPLYTLMKOVAGAG 309
Db 1173 DAGGLKTTTKTEGNNVVEDVDNDVGTGKATESYQYDGN-----VTSYKDAYGT- 1225
Qy 310 QPAILAEYSYSHNYVGGSGNGIMNNLNLGYLMTENYNGSTESRRYKQKEGHOIVRI 369
Db 1226 -----ETFEYKNNNDV-----TKMKDEGNVTDIAYDGLDAVBETDQSGKSSAAV 1271
Qy 370 ERTYNNYHLTSEKQONGYIQTETAYVAILIGNPFSQPSQPOLPTKTEYTRSDNSY 429
Db 1272 YDKRGNIQSSKLSASTNLIK-----DGSFEAQKSGNNLTASKD----- 1311
Qy 430 RSEITETTPDESQNPFLKVIKDKKTOKIISPS-----THMEYYPAGEVNDCCPEPYGPT 485
Db 1312 RKKIS-VIADKSG-----VLGSKALEVLSOSTSAGTDHGYSSATQVLEPNTTYL 1365
Qy 486 FVKKI---QTPYSEKDDPEKFIQY---RSLIGSOHVTLKIERHYSATQLNS-- 537
Db 1366 KITDLAKSAYFNIDLRDKDQKRIQWINEYSALAKNDWT-----KQITFTTPAAGK 1421
Qy 538 ---TLFOYNTDKSELGL-LKQTECTGKNGKTYSVHAKTYTKODDTLOOS-HSITTHD 592
Db 1422 AVYMEVDHDKDKGKAMPEDEVOLEKEGVSSSNPQNSFTSATENMNVSGASVDSSE 1481
Qy 593 NPTIHSQVRSRYTGRLFSDTDKDIYQMSYDGLRLTRT-----L 635
Db 1482 GFNDVDSLKAARTSASQASVTKQTVLQSGANDKPVYLTITGMSKASVYFTDEKDYSL 1541
Qy 636 NSGTPYANTLTYDEL---NNLQDNRPPFYI----- 664
Db 1542 QANVTYADSGTGYNAKFPSTQEMNAAVVIPTKPIINKVDISILFQKSATGTVWDDI 1601
Qy 665 -----TTTVDVNGQLNEPDGAGRHSQCLKSDGDKYTTHTQOYDEQGHNT 714
Db 1602 RLIEGSLTSTYSNNGNYVKEDEBEGYATIS---TYDETGK---KTSETAKGKXT 1654
Qy 715 STY--SDYLTNGRQQTQPDKVHLSMSKSYDNWG-QIANT-----HMSYGSSEKI-- 760
Db 1655 YTTDOAQLTNMTLSNGSTSIH---SYDEKGEVSKTIRAGADQTYKEBYDMGLV 1709
Qy 761 TVDPITLTATKQLOSNNNVOT---GKEVYTYTPSOQPIQITLFDRAHQLQSCHTLTRD 816
Db 1710 TTPDLGNVLASEYDANGLTFTISPNQNEV-----SLSYD 1744
Qy 817 GMDRVRKETDAIGCCTIYQYDNVNRVYQITLPDGTIVNRKAYPSTDTLITDIRVNGISL 876
Db 1745 GTRVKSYSYNGEYKIFYDYDKNGN-----ETSVNKEQN-----TT 1781
Qy 877 GQGFQDLSLTQSDGGRWAWYTYASAGNOCPSVTITPDSQFIHYQYQELDAVLQVA 936
Db 1782 KKRFTGKRNKRLTETLDRGSGQWTWYPSDCLKTF-----SWH-----G 1821
Qy 937 SNETIQQFSYNPVYGLLKAVAEQSUTPIYYPGRLMENINDMKMSYMLTLRGLENG 996
Db 1822 DQKGTQNFYN-----KLDQMIEKOSTSYSYSDYEN- 1854
Qy 997 YTDLTGTIQKISRDTGRVTOIKDSSIKTTLNVDLNRHIGSQVTDLATGMLTTTYEFD 1056
Db 1855 -----GNVQ-----TFITNGCGTSSFSYDERNLVLSIHLGDKNGGDIILTESYEV- 1898
Qy 1057 GUNREIRKLCDSGHTLIDIQSWLKTQQLANRIVKNGVLRQREOYSYDSRNLNLYKC 1116
Db 1899 -----DANGNRTTINS-----ASGVK---QYBYGKLNOLV-- 1927
Qy 1117 DGAECPDYKGVSHVTONFTYDIQNTIACHTTEADGTED--HATFFRANPTDCCQLETV 1174
Db 1928 -----ETHDEGTVA---EYTYDGFGRNKTV-TTIKDSSTKVANSFNIMN-----QJLKV 1973

```

```

Qy 1175 H-----HTHP-DMPDNI-----RLKYDKAGRVINITDNG 1203
Db 1974 NDESISYDKNGRNTSGKTYTTWDAEDNLTAIVKKGEDRPATYKIDENKRNLOKTVN-G 2032
Qy 1204 NTEFTYDTLGRLONGQSGVGYDPLNRVLSOKTDTLDELY-YRETMVNEVRNGEMIR 1262
Db 2033 KVTNIFYDG-----DSLNVLYETDADNNVTKSYTGD-----SQLLS 2070
Qy 1263 LRTGETIILAQASAVLLTGTDSQOSVILTSKQLSQEAAYAKG-HKSTANDA---S 1318
Db 2071 YTENGGKYFFHYVAHDIILASDSTGKTV-----AKQYDAMGNPTKTESDEVKDN 2122
Qy 1319 ILGNGERADPVSGVTHLNGYRSYDPTLMRFHT--PDSLSPGAGGINPYCYCLGPIN 1376
Db 2123 RRYAGYQYDEETGLYLMARY--YEPKNGVFLSLDPDSDSDSDLDQNGYAGNNPVM 2180
Qy 1377 RSDPSGHLSSQAWTGIQMGIALLLTATGMAIAAG---GIAAIASTSTTALAFA 1432
Db 2181 NVDPDGH--W-VMLVYNAGFA-----AYDGYKAYSGKGMKGAAMAAASNFGPKIFG 2231
Qy 1433 LSVTSDITSIVSGALDADSPKASSILGWVSMG---GAAGLAESAIKGTAKLATHGAF 1488
Db 2232 ASRAYFTK-----KAVKITGHTHGLNOSIGRNG-----GRGVNLRADKANA- 2273
Qy 1489 AEDGENALLKSTSESRIRKMGVTRSLDRE--IYRNEGQYIKDHSRGYTDNFMGKQBAI 1546
Db 2274 -----VASPKVTKOPNGATKYVGKATVVLNRGKVTIYV-----SS 2312
Qy 1547 LVHGDQKGFLYHTEGK 1563
Db 2313 RAKGSKHVFHTHGK 2329

```

```

RESULT 4
S38239
hypothetical protein - Coxiella burnetii
C:Species: Coxiella burnetii
C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Jul-2004
C/Accession: S38239
R:Thiele, D.; Willems, H.; Haas, M.; Krause, H.
submitted to the EMBL Data Library, October 1993
A:Reference number: S38215
A:Accession: S38239
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-336 <THI>
A:Cross-references: UNIPROT:Q45946; UNIPARC:UPI0000086171; EMBL:X75356; NID:G407370; PIDD

```

```

Query Match 4.8%; Score 429.5; DB 2; Length 336;
Best Local Similarity 34.3%; Pred. No. 2e-16;
Matches 121; Conservative 48; Mismatches 91; Indels 93; Gaps 13;

Qy 1348 MRFTPDLSLSPFAGGINPYCYCLGPIRNSDPSGHLSSQAWTGIQMGIALLLTATG 1407
Db 1 MRINCDSWSPFAGGINPYCYCLGPIRNSDPSGHLSSQAWTGIQMGIALLLTATG 60
Qy 1408 MAIAAGIAAIASTTALAFALSVTSDITSIVSGALEDASPKASSTLIGVWSMGMA 1467
Db 61 TSAAGAIASALIASAIAISLVVGTGLVADVASIASGALEDANPQASATLGMISLIG 120
Qy 1468 AGLAE--SAIKGTGLATHGAFEDG-----ENALLKSTSESRIRKMGV--RS 1513
Db 121 PGAVSLATYARAGKGL--ISGLAKGGKIRSQSPVQGISYSLSGDPLRGGPFPQS 177
Qy 1514 LDREIVANEE-----GOVTKDHSRGYTDNF-----MGK 1541
Db 178 LSRVYTAPEBMRPAGLNYMHKIVQKSLGYQHFGADREIFGYEIRPIEFPFRRSBITK 237
Qy 1542 GQQAII-----VHGDQKGFLYHTEG-----NKNNGKGPYTRHPTPEQLVDY 1561
Db 238 RDIIVILSGTHGRVHGDN---WTSQGLRRPDILERAFFYIEDVQNYKG-----QLNGR 285

```

Query 1582 LKDNINVDLTCGG-----DKPVLHLSCKGSSGADKKAKYIN-RPVLAY 1625  
Db 286 VKIYDAMGMITSESEGRVVRNNGHVLIGCKMGRDEA---LTHRNLRPVISTY 335  
RESULT 5  
AF1489  
cell wall-associated protein precursor wapa (B. subtilis) homolog lin0454 [imported] - L  
C;Species: Listeria innocua  
C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 09-Jul-2004  
C;Accession: AF1489  
R;Glaser, P.; Farnaul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duseburg, O.; Entlian, K.D.; Fahl, H.  
D.; Jones, L.M.; Kargel, U.  
Science 294, 849-852, 2001  
A;Authors: Krefel, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schlueder, T.; Simoes, N.; Tiberre, A.; Vazquez-Boland, J.A.; Voss, H.; Weiland,  
A.;Title: Comparative genomes of Listeria species.  
A;Reference number: UNIPROT:Q92EK5; UNIPARC:UP100000CC234; GB:AL592022; PIDN:CAC95686.1;  
A;Accession: AF1489  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-2167 <GLA>  
A;Cross-references: UNIPROT:Q92EK5; UNIPARC:UP100000CC234; GB:AL592022; PIDN:CAC95686.1;  
A;Experimental source: strain Clijp1262  
C;Genetics:  
A;Gene: lin0454  
Query Match 4.3%; Score 386; DB 2; Length 2167;  
Best Local Similarity 19.4%; Pred. No. 1.3e-12;  
Matches 327; Conservative 212; Mismatches 582; Indels 568; Gaps 67;  
Query 17 NNEFTQANNTSAVSGVDPRGLY---NIQITIGHI-VNGNGL-----GP 59  
Db 718 NNTTSVSEPRPTSPREALNRQLGKVDYWTSPVRGSEVNAATNGFLFHEHEDPMLBGRGP 777  
Query 60 TLPLTSLSPKLTIDIGFIGNFGLSYDRKNSLSLSTGENYKVIETDKTV----- 112  
Db 778 SINNRFTNSQDDATGIFGKMTSTL-----EKLVEENG-NLWVWSDKKIRFTKGG 831  
Query 113 -----KLQOKLKD-----NLFKDKLXENYRIIHKSGDIEV 144  
Db 832 DKYAPPGIYSEITKADGVLKIEEDKSETRFLVDGRKSEKDTGKMLTYEYTDGKLTG 891  
Query 145 LTGNNNNAFDL---KVPKLLNPAGHAIFYIDMNPENATQPLNRYDLDHDDPLNLTLE 200  
Db 892 LRDSGRTVTLTYEGELVKEIVGPEDKISYTN-----D 926  
Query 201 YQGLIKTILTLFPGQKEGYRTLEPLNRQLNSIHNFSLGNEPPLTWSFGYTPIGKNGILG 260  
Db 927 KQELISSSTARGLKYRYGYTDGL-----LTSIYDPKHTBEKPYATTFAVE----- 972  
Query 261 QMITSMTAPGGLKETVNSNNNG-----HHFPOSANLPVLPTYTLAKQVGA 308  
Db 973 EKLTEITDPVQKKTTLSDYKAEQQTTLTNEKKKKTIVSYNDAGN-----PKKEIYDA 1024  
Query 309 GQPAIQAEYSYTHYVGGSGNGINNNKLDNLGLMTEYNGSTESRRYKXKEG----- 363  
Db 1025 DGLKLTITTYTESNN-----LVKAVNPKGQSEYTAIYADGNITAT 1065  
Query 364 DOIVRIERTYNNYHLTSECKQONGYIQTETAYVAIIGHNFDQSPQFOLPKTETETWR 423  
Db 1066 DAYGIESYTYNDNNVTS-----STDTB-----GRKTVTYD 1097  
Query 424 SADNSYSEITE-----TTPDESGNPL-----TKYIK 450  
Db 1098 GADAVSETLATESQVSVYQDAYGNPIRGSELSSGNLQNSGFEGAGVSNMTLLQS 1157  
Query 451 DKK-----TG-----KISP--STHMEYPPAGEVONCPPEPYGFTFRFYK 489  
Db 1158 DAKSGMTDNTQSAFGALGGSGSVKLTSEANSTVKGYSSTVORAVVEBETTYTFSAWK 1217  
Query 490 -----IIGTPYDSEFKDPEKFIQYRSLIGSQSHVTLKIEER-HYSATQLLNSTL 539

Db 1218 SGMTNADALLIGRLQDANAKVDYDAGVW-----OSNPAISIKKQGDVVKQL----- 1264  
Query 540 FQNTDKSELGRLL-----KQTECTGENKKTYSVNAKFTYTKQDDT 581  
Db 1265 -TERTSKNTRQVLLYLDNEQAPRHKGTIWDNVQPEKGSVASSINPVNNSENNNGT 1323  
Query 582 LQOSSHSTTHDNFT---IHRSQVRSRYRGLFSDTDTKQIVTQMSYD-----KLGRLLTR 633  
Db 1324 LPTGMRTGNTATLQAKVVDNNGSHSGDSANVFFRKATSEAYTHIYQDVPVNGEAKAL 1383  
Query 634 TLNSGTPYAN-----TLTYDEL-----NNLQ-----DNRPPVITTTD 668  
Db 1384 SALSKEADAKANGSVATMSNDYSVWGTVVYQDPTSSVQCPPLGTNDMNRSAVVVAPTK 1443  
Query 669 -----VNGNQL-RNEPDGAGRHSQGLKQSDGKREYTI 701  
Db 1444 PVKAIKYTTMFRNGLTGKAMPDVRVYEGVLITNEVDASGNTV----- 1487  
Query 702 HTOQYDEGHRHTSTYSDDLTLNGRQOTDPDKVHLSMSKSYDNMGQIANTHWSYVSEKIT 761  
Db 1488 -TASYDEGRKISFTYDIY---GNKTSIEDKKNKKTITLTDADNALIDTLANGSTVAVK 1543  
Query 762 VDPITTLTATQLOSNMNVOTGKEVTTTPSQPI-----QITLPEAGHLQSCHTLT 814  
Db 1544 YD-----DNGNTTEKNVTASGKTQKUIVEYDVNDKITAFTDA-----LNRTI 1585  
Query 815 RDGMDRVKRETDAL---GOCITIOYDNNRVIOITLPDGTIVNRKVAFFSDTLITDIRV 871  
Db 1586 KYEDDAAGNETKAKMENGRTVESTYDSADMDGIMKADKAFKQYQPNNGQTYVTD-EI 1644  
Query 872 NGISLGOQTFDGLSRLTOSQDGRWAAVTVYSAQNDQPSFTVTPDQGFHYHQYQPELD-- 929  
Db 1645 NSI-VTDKTYDANRIKVAERGDVSYT-----YKDKFTYDNK 1682  
Query 930 ---DAVLQVANSR---ITQGFSTNPVTGALLKRAVAGGSLTPYPSGRLKAMENIDMK 982  
Db 1683 GKTDKGEVAIINMGDYTAKTSTYTNLD-----RNRVNDGS 1719  
Query 983 KMSYL-WTLRGLNGYTDLTGTOKISRDNHGRVTOIKOSSIKTTLNVYD-LNRHISQV 1040  
Db 1720 KNAIFEDFEGENIVYTAGNGTAAANYTYDSTQKTAIAISASQTULDEYTYDDAASNR 1779  
Query 1041 TDLATGMLTTFVFEDELNREIGRKLCDSSGHTLDDIOSWLKTOQLANRYKLVGLVORT 1100  
Db 1780 TSDINKQDKTTFYDAVNO-----LTKETLPDGYTK----- 1811  
Query 1101 EQISYDSRNRLNQKCDGACPT--DKY--GHSIVTON---FTYDIIGNITACHTPADG 1153  
Db 1812 -ATYUDGFGRNTOVALISGSETKITDASYNQNLVSNMGALTYDANGNRT-----SDG 1864  
Query 1154 TEDHATFKANPDPQCLTEVHNTHPMPDIRLKYDKAGAVINITDNHGTENFTYDTL 1213  
Db 1865 KYTYT-----WTDGRLSIT-KKGSSEPT----- 1889  
Query 1214 GRLONGGSYVDPLNRVLSQKTDLDCELY--RETMLVNEVRNGEMIR-LLRIGE 1268  
Db 1890 -----SYTYDDNRRLSKTYDVGTVYHNGDSIDLXYETDQGVKVVQYVYSDDN 1940  
Query 1269 TTIAGQASAKULLTGTDSQSVILTSDKQ--LSQEAYSAGK-----HKSTANDAS 1318  
Db 1941 VRLAMKNGKTLVYHNAHGDIVALTDEAGKIYAEYVADAGNVLKNTASTEEKAKANYG 2000  
Query 1319 ILGNGERA-----DPVSGVHNLGNGYSYDPFLMRFTPDPDLSFPGAGGINPY 1367  
Db 2001 YAGTYTDEIJBQYLLMAYIEPEGV-----FAYID-----YGBDDP--QTMNGY 2046  
Query 1368 SYCLDPIRNSDPSGHLSSW---QAMTIGMGIAIGLLTIATGMAIAAAGGIAAIASTAS 1424  
Db 2047 NYANNNPVMWMDPDGNAVAMWIAAGYGAFBEGAEVLLITTKKQNMK-----GFGKAVVGA 2101  
Query 1425 TTALAFGL 1433

Db 2102 VLGI6GFKL 2110

## RESULT 6

T37218

hypochemical protein SC2H4.02 - Streptomyces coelicolor

C/Species: Streptomyces coelicolor

C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004

C/Accession: T37218

R/Oliver, K.; Harris, D.; Parkhill, J.; Barrell, B.G.; Raftery, M.A.

submitted to the EMBL Data Library, September 1998

A/Reference number: 221615

A/Accession: T37218

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Features: 1.2183 &lt;OLI&gt;

A/Cross-references: UNIPROT:O86585; UNIPARC:UPI00000DADB3; EMBL:AL031514; PIDN:CAA20596.

A/Experimental source: strain A3(2)

C/Genetic:

A/Genes: SCOE2H4.02

Query Match 4.3%; Score 381; DB 2; Length 2183;

Best Local Similarity 19.8%; Pred. No. 2.5e-12;

Matches 366; Conservative 221; Mismatches 669; Indels 594; Gaps 91;

99 TGENYKVIETDKYKLOOKK-----LDNLFPEKDKENYRIHKSGLIEVLTGF 148

Db 620 TGS-----TDOSLMDLKHGKGTDLSDAVKFSHVMPN--RVDGSPDDI----- 666

Qy 149 NNNAFDLKVPKLLNPAHAIYIDW-----NFEATOPRLNRIYDDLGDHDIPL 196

Db 667 --LSFERPRARTVSEGAQITVDYLRADCAVQTMKADENTRCRFPVWSPYQGEPI 724

Qy 197 LNEIYQGLITLILFP-GQKEGYTELRLNROLNINHPISG-----NENPL----- 244

Db 725 LDMFOKRPVSVRTTDLGSEA-----VQHTVEYSGGAMHVNDPLTPAKER 773

Qy 245 TMSFGYTRIPKNGILGOW-----ITSMTAPG--LKEVTNYSNNNGHHPQSANLPL 296

Db 774 TMS-----OMRGYKVTHTYTPSGGTRATTVTYLRGMDGRVLDGDKTPD 820

Qy 297 PYUTLMKQVPGAGPAI-----QAEYSYSHNYVGGSGNGIWNKLDMLYGLMTE 346

Db 821 PDARRKAEVSGVTAATDSDQLAGPQESVAVYGDGEGVTVNDPMSKST----- 871

Qy 347 YNTGSTRRYKQKEGHDQVRIERTYNNTHLTSCKQONGY--IOTTETAY--YAIIG 402

Db 872 ----ATQKSYADTEAY--YVRVAGS--HARTITSRLNPDYRVTRTGYDDYGMV 921

Qy 403 HNPDSOSQFOLPRTKETWRSADNS-----YRSEITETPD--ESGNPL 445

Db 922 SVELDGDSVTDGCKCTRTYARNDAAGLTALVSRTRIVGRACSVTSDALDPAARF- 980

Qy 446 TKYLKDKTKOKIISPTNME--YYPAGEVNDCEPEYGFTRFKKIIQTPYSEFPDOP 503

Db 981 GDVVSDDAT---AVDTTMSATQKPTKGA-RMSGRKAGYGGDRPLMOTTAATYDY-DTL 1035

Qy 504 EKPIQVRYSLIGSOSHTLKIIEBHYSATOLNLSLFOYNTKSELGRLLKQTECTGEN 563

Db 1036 GRPQVQV---NTDDVITSKTEVQPPAAGPLTSTTV-----N 1069

Qy 564 GKTVSVNKKFTYKODDTLQOSHITTHDFTNRSQVRSRYGRLESFDPDTKDITQMS 623

Db 1070 AK-----GHRTTVKDFAL-----GADLKVDANGKVTESA 1100

Qy 624 YDKLGRLLTRTL-NSGTPYANTLLVDEYELNNLDONDPFVITTTDVNGQLRPFAG 682

Db 1101 YDLSGRATSVWLPKRSALGKTANYVYGV--VASTSLPWSSAT-LNG-----DGG 1150

Qy 683 RHVSQCKSDGDKFTYTHITQYDEGGRHHTSTYSYLTNGR----- 725

Db 1151 YRTTYEYDS-----LRTTRQVAPSAQGRVIAQTLYDGRGLPYTAQADIWDPAP 1203

Qy 726 ----QOTDPKVALSNKSYDNMGOIAN--THMSYGVSE----- 758

Db 1204 AGKIVQIDGGQARQDSDSYVDGWRVTKAVTYSYGTENAVDTRAGDVLITGAPBGSA 1263

Qy 759 -KITVDPITLTATKQLOSNNSNVQGEVTTYTPSOQPIQITLPDAGHLQSCHTLTRDG 817

Db 1264 NAVVTAFGTRVRRVAGTQAPAGTYMTTRVA-----FDADROKSTIADNRSA 1313

Qy 818 W-----DRVAKETDAIGQCTIYQDNNRNIQITLPD 849

Db 1314 WYTYTDLFGQVSVTPDCKTTEYDALBRAVKSIDGRSEVLLFEYDVGKTKGMQSA 1373

Qy 850 GTVNRKRYAPFSTDTLI-----TDIRVNGISLQO-----QTPDGLSRLTQSO----- 891

Db 1374 KTGAN-KLAMSFDTLAKQODPAVYEGSETRATQKTRRDPLKYKTNNLPLPAND 1432

Qy 892 --DGRVMAVYTSAG-----NDQCPSTVTTPDQFTHYQYQPELDAVLOVASNEIT 941

Db 1433 PLVAAQVPAALASTGTNLDTGVKQAAAPAVAGLSAETVSYYT-----DELQGVLTAKGT 1487

Qy 942 ----QPESTNPTVGAALLKAVAEQSLTPITYPSGRKLMENINDMKKGSYLTMLRGLNGY 997

Db 1488 TGYLQQAAYS-----LGLRQMTLATDPGAKKVIYLNND--YAGTRRLTRSY 1534

Qy 998 T--DLGTI---OKISRDTHGRVTOIKDSIKTTLNVDDLNRHIGSQVTLATGHMLTTT 1052

Db 1535 VYDDVHGFMLQELAKYQDDAGNITSVDA---TLTG-----GTGKADHOC 1576

Qy 1053 VEPDGLNREIGRL-----CDSSGHTLDIQOSMLKTQOLANRIVKLVGLORTBOYSYD 1106

Db 1577 FTVDG-HRRLSEAMTPETADCSGRT-----VAGIGGAAPWTSYQYD 1619

Qy 1107 S---RRRLNOKYKDGACPTD-KYG-----HSI---VT-----QNTFYDIYGNITACHTT 1149

Db 1620 DSGIRSKQTEHRSQDVTTEYEGTAEGPRLSATVTAENASYYTDDTG----- 1672

Qy 1150 PADGTEDHATFKPAPLTPCOLTEVHHTHPDMDNIRLKTDKAGRVNITDNNGTENFT 1209

Db 1673 -----TE---TRGVAAQTQILDMNAGRLAGVSEPPAAGKPRAT 1707

Qy 1210 YDTLGRLONGQSGVGYDPLNRLVSOQKTDLDC--LYRETMVLNVEVR-NGEM-----I 1261

Db 1708 -----GTAAYYDAGDILLIRPPTTDETVLYLGTETVHLKVSNGCAKALSGA 1756

Qy 1262 RLIRTGETTIAOQASAKVLLTGDSQSVLITSDKONLSQEAISA-----YGR 1309

Db 1757 RYKAGSAVIAVYTS---AGVSGTKLTFPLADHGTSGLAINDTLAPAKMSTPFGA 1812

Qy 1310 HKSTANDA--SLGVNGERADPVSGVTHLNGVRSYDPTLMRPHTPDS--LSPGAGGINP 1366

Db 1813 PRGTAGAMPDDRGFLGKPADAAATGLTOL--GARQYDPDGRFLSYDPLLEPKPTLNG 1870

Qy 1367 YSYCLADPPIRSDPSGHLISQWAMTGLIGMIGLTLAT--GGMAIAAGIAAIAASTS 1424

Db 1871 YAVASNSPVTNSDPG-----TSDGLG--GLLGAIGALIGGVAVIGAVGAAI--TA 1919

Qy 1425 TTLALFALSVTSDITISYSGALEDASPKKASITLGVNSMGGAAGLAEKIKGTKLATH 1484

Db 1920 VGSILGGGG-----GGGWGTAAPTSSG--GW-----TQPLTKQMTPGATY 1956

Qy 1485 LGAPADGENALLKS-----TSES-----SRIKGVTRSLDREIVNER----- 1523

Db 1957 -----NFTKSMDLPRNPSPQSLIEMLAMPDQIAYS--DPRKAANWETSRLIFG 2005

Qy 1524 ----QGVINDHS--RGYTDNFMGKGBQALIVHGDK-DEGLYHTEGKNHG-KQPYTRAT 1574

Db 2006 WLMGGGYPLEHEDFRG-----GDAFTSLADDETISGLRSMVQGARKGTKGAPAKEV 2059

Qy 1575 PEQULYDKNNITVDLTGGDKFVHLLSCYKSSGGAADRAKAKIINRPV----- 1623

Db 2060 GFQYVD-----EGPEGSPWYKNSLGAAGAKIDAGVLTNGAVGTGQADAFIG 2107

Qy 1624 AYSNKPTISGLARIBRKDPFLKSTYHSYDPRKILIGRTEKTVKPTFRP 1673



Matches 223; Conservative 121; Mismatches 308; Indels 327; Gaps 48;

QY 542 YNTDKSLGR---LTKOTBCTKBNKKTYSVHAKFYTKDDTLQOQSHSTTHDNFTIH 597  
 DB 32 HHTDSGHOYRLDMNLALERSLCTVDSMGT-----SCHMMDAOGLVAY 74  
 QY 598 RSQVRSRYTGR-----LFSDTDKDITVQMSYDKLGRLLTTLTANSGTPYANTLTVDYE 650  
 DB 75 RDEAGCMTTRMSDEERLLGMTDAOQGMKRYVDRGLH-----TETHD-P 120  
 QY 651 LNNIQQDNNRPF---VITTDVNGNQLRNEFD-----GAGHVSQCL 689  
 DB 121 LGREYQOMPHVHMHPETEVDAAAGVARYEDERGNLAQVSDPLHQTAVGVYDHH--GQVV 179  
 QY 690 K--DSDDGKRYTHTHQYDEG---RH-----HTSYSDYLNGNQDTPDKVHLSMS 738  
 DB 180 RITPARGDKYL-----QWNEEQGLMHTTDCSGQTMFEDERLLEVTDAES--NSTR 232  
 QY 739 KSYDNMGQIANTHMSYSEKITVDPITLTATKOLQSNMNVQTEKVTYTPSQOPIQI 798  
 DB 233 YSYDGNHLEVMADRTERTYQD-----AAGRLVKTSPAG---QI 272  
 QY 799 TLEFEGHLOSCHTLTRDGMVREKETDAIGQCTTYQYDNYRVIQITLPDGTIVNRKYA 858  
 DB 273 TRMQ-----RDGQGRVRRQTDATGRATAYEDAGRLTTLTNGESYRRRY- 319  
 QY 859 PRSTDTLTDRVNGISLGQTPGSLRSLTQSP--GGRVMAVYTSAGNDQCP----- 909  
 DB 320 -----VLDKRVTEQTPDGGRRRYGVNALAVAVIYGBERG 356  
 QY 910 -----STVTPDGOFIHYQYQPLDDAVLQY-----ASNEITQOF 944  
 DB 357 GEIRHGLEBDAAGRLTAKITPE---TREYRYDADRLERRRRHAAGEGEVLR-RF 412  
 QY 945 SYNEVTGALLKAVAG-----QSLTPYYSGRKLMENINDMKMSYLTWLTGLEN 995  
 DB 413 SYDSAGNLSHETAGVQLHRYDVQGNRTETQMPDGR-----TLRYLYX 456  
 QY 996 GYTDLTGTIOKISDTHGRVQIIDSSTKITLANTDNLNIGSQVDTLATHMLTTVEF 1055  
 DB 457 G---SGHLQOI-----NLGRDVISSETR----- 476  
 QY 1056 DGLREIGRKLCDSGHTLDIQSMWLTQOLANRIV--KLNGVLQRT---EQYVDSNR 1110  
 DB 477 DHLREYQV---SQGR-LDMRMYDRGRRLTKLTKGMMGVPEFTIDREYVSGQDE 531  
 QY 1111 LMQYKCDGAECPDKYGHISIVTONFTYDIYGNITAC--HTTPADGTEDHATKFPANPTDP 1169  
 DB 532 LLK-----KRHSQGVTDYF--YDTGRTITACRNEAVLDSMQVDA--ANLDRR 577  
 QY 1170 Q-----LTVNHTHPMPDNIRLKYDKAGVINITDNGHNTENFTDTL 1213  
 DB 578 QGETAAGAGSVDFNRITSYRGLH-----YRYDEYGRVVERKRGNG--TOHYRMDAE 628  
 QY 1214 GRL-----QNGQGSVYG--YDPLNRLVSOQTDITLCELYRETMLVNEVNGEMIRLLR 1265  
 DB 629 HRLREAVNTRGYRRRGYVYDAPGRVREKHELDAGEKPYRTTFLMDGMLAQBCHLR 688  
 QY 1266 TGETTIAQOR-----ASKVLTLGTD--SQQSIVLTSBKONLSQEA--YSAY 1307  
 DB 689 SSSLVYISDRGSHPELARVDBAPAGEADEVLYYHTDVNGAPEEMTDGGGIVMEAGVQVW 748  
 QY 1308 GK---HSTANDASILGNGERAPVSGVTLGNGRYSDPTLKRFRHTPDSLSPFG--AGG 1363  
 DB 749 GNLTHEKSTRPVQGNLRFQGYLDRETGLAH--NLRYFYPDIGKIFISGD---PIGLAGG 803  
 QY 1364 INPYSYCLGDPINRSDPG 1382  
 DB 804 INLYQYA-PNPLSYIDPLG 821

RESULT 9  
 C64805

thsc protein precursor [similarity] - Escherichia coli (strain K-12)

C:Species: Escherichia coli.  
 C:Date: 12-Sep-1997 #sequence, revision 17-Sep-1997 #text\_change 09-Jul-2004  
 C:Accession: C64805; 163400; T88912  
 R:Blacler, F.R.; Plunkett III, G.; Bloch, C.A.; Berna, N.T.; Burland, V.; Riley, M.; Co  
 A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A:Title: The complete genome sequence of Escherichia coli K-12.  
 A:Reference number: A64720; MUID:97426617; PMID:9278503  
 A:Accession: C64805  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-1397 <BLAT>  
 A:Cross-references: UNIPROT:P16918; UNIPARC:UPI00001338AF; GB:A5000173; GB:U00096; NID:G  
 R:Feulner, G.; Gray, J.A.; Kirchmann, J.A.; Lehner, A.F.; Sedocky, A.B.; Vlazny, D.A.;  
 U. Bacteriol. 172, 446-456, 1990  
 A:Title: Structure of the rhaA locus from Escherichia coli K-12 and comparison of rhaA w  
 A:Reference number: I54935; MUID:90094253; PMID:2403547  
 A:Accession: 169400  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1221-1397 <RES>  
 A:Cross-references: UNIPARC:UPI000017888F; GB:M29718  
 A:Note: the sequence is revised in GenBank entry ECOLRSCA, release 114, (PIDN:AAC63073.1  
 R:Ohshima, T.; Alba, H.; Baba, T.; Fujita, K.; Hayashi, K.; Honjo, A.; Ikemoto, K.; Inada  
 T.; Mizobuchi, K.; Mori, H.; Motomura, K.; Nakamura, Y.; Nishimoto, H.; Nishio, Y.; Sai  
 DNA Res. 3, 137-155, 1996  
 A:Title: A 718-kb DNA Sequence of Escherichia coli K-12 Genome Corresponding to the 12.7  
 A:Reference number: Z25006; MUID:97061202; PMID:8905232  
 A:Accession: T48912  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1397 <OSH>  
 A:Cross-references: UNIPARC:UPI00001338AF; EMBL:D90709; NID:G1651305; PIDN:BA35359.1; P  
 A:Experimental source: strain K12; Kohara clone 174  
 C:Genetics:  
 A:Gene: rhaC  
 A:Map position: 15.7-16.0  
 C:Superfamily: rhaF protein

Query Match 3.9%; Score 342; DB 2; Length 1397;  
 Best Local Similarity 21.4%; Pred. No. 1,7e-10;  
 Matches 256; Conservative 157; Mismatches 382; Indels 404; Gaps 63;

QY 418 KTEW-RSADNSYRSKSTETTFPDSGNPLTKV-----KDKTKIIS-----PS----- 461  
 DB 218 RTQTFHREAAGEFSGEITGV--DAGRHFRVLVLTQARABEARQQAISGTEPSAPEDT 276  
 QY 462 -----THMEYRPAQGVNCPPEP---YGF-----RFVK 488  
 DB 277 LPEYTERGRDNGIRLSAVMLTHDEYF---EULPAPALVRYGMYTRGELAAVYDSNT 331  
 QY 489 KLIQTPYDSEK-----DDEKEFYQYVSLIGSOHVTLLKIEBRHSATOLLNSTL 539  
 DB 332 QVRSFTYDDKRGWVAHRTGRPE--ICRYD--SDGRVTEGLNAGLSY----- 379  
 QY 540 FOYNTKSEGLRLKQTEC--TKGNGKTYSVHAKFYTKDDTLQOQSHSTTHDNFTIH 597  
 DB 380 YQEKDRITTTDSLNRRREVLHTQEGG-LKRVVK-----EHADSVT 421  
 QY 598 RSQVRSRYTGRLFSDDTKOIVTQMSYDKLGRLLTTLTANSGTPYANTLTVDYELNQLDD 657  
 DB 422 QSQFDA--VGRKRAQTDAGRTTYSFDPVVTGLTIRLT--TPGGRSAFY----- 468  
 QY 658 NRPFVITTTDVNGNQLRNEFDAGRHVSQCLKSD-----GDGKFTYHTQYDEQGRH 712  
 DB 469 NHHSQLSATGPDLRIIRREYDEWGRLLIGTAPGDITRRYDNPBHDLPACATEDATGSR 528  
 QY 713 HTSTYSYDLTNGRQCTPDKVHLSMSKSYDNMGQIANTHMSYSEKITVDPIT-TLVATK 771  
 DB 529 KVTWTSRY-----GQLSFTDCSGVYTRVYDHRFGQVTAVHREBGLSOYRAYDSRQGLIAVK 585



```

Qy 772 QLAGSNMNVQTKKEVITTPSQOPIQITLFDAGHLQSCHTLFDGMDRVKETAIGQ- 830
Db 586 DTQGHETRYE-----YNAAGDLTT--VIAPDG--SRNGTQYDAMGKA 623
Qy 831 -CTI-----YQYDNKRVYQITLPDGTIVNRKXKAPSTDTTLITDRVNGISLQQT- 881
Db 624 ICTTQGGITRSMEDVAGRVIRLTSNGS-----HTFRYDVLRLIQETGPDRTQRYH 678
Qy 882 -DGLSRLTQSGDGRVMAVYTSAGNDQCPSTVITPDGFIHYQYQPELDADVLANVASEI 940
Db 679 HDLTGKLIRSEDEGLV-----THMYD--EADRLTHRTVANGT 714
Qy 941 TQOQSYNPNVTAL--LKAVAEGSLTPY-YP-GRLEKMEI-----NDM---KKMSY 986
Db 715 AERMOYDE-RGWLMDISHISEGHRVTYHYGDSKGLASEHLVYHNPQTNELMQHETRH 773
Qy 987 LMTIRGLENGTDLTGITQKISRDTHGAVTQKSSIKTITLNYDLNRHISQVTLATG 1046
Db 774 AYNAQGLAN--RCIPDSLPVEMWLYG-----SGWLKMGKLG 808
Qy 1047 HMLTTPYVF--DGLNRBEIGRKLCDSSGHTLDIQOSWLTQOLANRIVKLANGVLRTEQYS 1104
Db 809 D--TPLYVETDRLHRETLR-----SGRYELTAYTPAGOLQSG--HLNLSLG-DRDYT 858
Qy 1105 YDSNRRLNQYKCDGACPTDKXHSIVTQNFYDIYGNITACHTTFAD-----GTEDHA 1158
Db 859 WNDNGELIR-----ISSPRQ-----TRSYSTVTRGLTGVTAAANLDIRIPYTDPA 906
Qy 1159 TPKRANPPTDPCQLEVNHTTHD-----MPDN-----IRLKYDKAGVINIITD- 1200
Db 907 GNRLEDP-----ELHPDSALSMWPNRIADAHVLYRDRHGRLTEKTLPIEGV 956
Qy 1201 ---NHGNTENFTYDTLGLQNGQSGSVG-----YDPLNRLVSOQTDLTDEEL- 1244
Db 957 IRTDDETRHRTYDSQHLVYTRQYABPLVESVYLDPLGRYAKAVMREREDLTGM 1016
Qy 1245 ---YY---RETMVNE-----VRNGEMIRLR--TGETIIAQORA-- 1276
Db 1017 SLRKPQVTWYGMDQDLTTIQNDRTRIQITVQPSFPLIRVETATGELAKTORSLAD 1076
Qy 1277 ---SKULLGTDSQGSVILTSKQNSQRA----- 1303
Db 1077 TLQSGGEDGGSVFPPLVQMLDRLSEELA--DRVSEBSRRLWASCGLTVAQMSQM 1133
Qy 1304 -----YSAYGKHSTANDAS-----ILGVNG 1324
Db 1134 DPVYTPARKILHYCHDHGRLPLALISTGTTAWIAYEDWGNLNEENPHOLQOLIRLP 1193
Qy 1325 ERADPVSGVTHLGNQYRSYDPTLNRFTTPDLSPPG--AGGINPYSCIGDPIINSDPSG 1382
Db 1194 QQYDEESGLYY--NHRRYDDPLQGRYITQD--FIGLKGWNPYQYPL-NFISNIDPLG 1246

RESULT 10
R90658
R90658 R90658 core protein with extension [imported] - Escherichia coli (strain O157:H7, substrain
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C:Accession: E90658
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasaawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A:Reference number: A9629; MUID:21156231; PMID:11258796
A:Accession: E90658
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-1404 <HAY>
A:Cross-references: UNIPROT:O8XED9; UNIPARC:UPI00001653A9; GB:BA000007; PIDN:BAB3660.1;
C:Genetics:
A:Gene: EC0237
C:Superfamily: rhes protein

```

```

Query Match 3.8%; Score 341.5; DB 2; Length 1404;
Best Local Similarity 20.2%; Pred. No. 1.8e-10;
Matches 255; Conservative 130; Mismatches 370; Indels 505; Gaps 59;

Qy 376 YHL-LTSECKQONGYIQTETAYAILGNFDSQPSOLFPT-KTETMSADNSYSEI 433
Db 245 FHLVLTQAOGRAVFRKQATSLSSPAGR--SASSLVPPDPLPACTEAGDANGISLEA 302
Qy 434 TETTPDES-----GNPLTVINDKTKQKISSTHMEYRPAGEV-----DNCPREYGF 483
Db 303 VMLTHDPAYDELPAADLARYT-----YTASGEMLAVYDRSGTQVRGF 345
Qy 484 TRPVKLIQTPYDS-----FKDDPEKFIQYRSLYLSGQSHVLTKEERYASATOL 534
Db 346 A-----YAEHAGRWVAHHYAGRPS--RYRRDDIG-----RVTEL 379
Qy 535 LN-----STLFQYNTDSEGLRLKQTEC--TKBENKTSYVNHKFTYTKODDTLQOSHST 588
Db 380 VNEGGLDYREYQDQRTVITDSINRREVLTYBEQGG-LKRVVVK----- 422
Qy 589 TTHDNFTIHSQVRSRRTGRLFSOTDTKDLVTQMSYDKLRLTLRLTNSGTPYANTLYD 648
Db 423 -EHAQSLTSE-----YDAERLKAQT----- 444
Qy 649 YEALNLQDDNRPPFVITTTDVNGNQLNEFDGAGRHSQCLKDSG-----DGKPYT 700
Db 445 -----DAAGRTEYSLHMAAGAVTAVTGPDGR--T 472
Qy 701 IHTQOYDEGRHHHTSYSDYLTNGROQTPDKVHLSMSKYDMWGGIANTHWSYASEKI 760
Db 473 VR--YGNVSQRQVTSVTPDGLRSSREYDEKGRILAETSRSGE-----TRYSYD- 520
Qy 761 TVDPITLTKAQLAGSNMNVQTKKEVITTPSQOPIQITLFDAGHLQSCHTLFDGMDR 820
Db 521 -DP-----ASELPTGIQDATGSTKQW-----AMSR 544
Qy 821 ---VRKETDAIGCCTIYQYDNVNRVQITLPDGTIVNRKXKAPSTDTTLITDRVNGISLG 877
Db 545 YGOLFTPTDQSGVTTREYDRYQGLAVHREBGSITSYSNP-----RG 588
Qy 878 QQTFDGLSRLTQSGDGRVMAVYTSAGNDQCPSTVITPDGFIHYQYQPELDA--VLQ 934
Db 589 Q-----LVQKDAQGRFTRFYESAAGDL--TAIVAPDSRSSEIQY-----DAMGRAVS 634
Qy 935 VASNEITQPSYNPNVTGALKKAVAGOSLTPYIPPSRLKMEININDKMSYLTLAGLE 994
Db 635 TTQGGITRSMGYDAGRITVLTNENGSQSTFRYDPVRLTEQRGFDGRTQRY----- 686
Qy 995 NGYTDLTGTT-----QKISRDTHGRTQIKDS-- 1022
Db 687 -HYDVLGKLTQSEDEGLVTLMYHDASDRITHTRVNGDPRAEQWYDEHGMVLTLSHNSBG 744
Qy 1023 -IKTTLYDNLNRHISQ--VTDLATGML----- 1049
Db 745 HRVSVHYGYDDKGLTGERQTVENPETGEMLMHEHTGAYSEQGLATROBEDGLPYEWM 804
Qy 1050 -----TTTVVF--DGLNRBEIGRKLCDSSGHTLDIQOSWLTQOLANRIVKL 1093
Db 805 TYSGGYLAGMKLGTFPLVEYWRDNLHREYASF--GGAVELLATWNTSQQLSRHLNL 861
Qy 1094 NGVLAQRTQYSYDSRNILNQYKCDGACPTDKXHSIVTQNFYDIYGNITACHTTFAD 1153
Db 862 P---QLDRDQDMNNGDL--IRISPG-----ESREYRSQDGRLLGVHTAANTL 906
Qy 1154 TED--HATFKRANP--TDPQCLEVNHTTHD-----MPDNIRL-----KYDKAGRYIN 1197
Db 907 DIDIPVATPDAGNRLPDP-----ELHPDSTLTPAMDN--RIADAHVYVRYDEYGRILAE 958
Qy 1198 ITD-----NHGNTENFTYDTLGLR-----QNGQGSV--YGVDP- 1229
Db 959 KTDRIPEGVIRMDERTHAYDSQHLVHTTRIQHGEPQVESRYLDPLGRRTGKRVWR 1018

```



Qy 1230 -----NRLVSOQKTDLDCELYR-----ETMLVNEVRNGEMI 1261  
 Db 1019 RERDLGTGMSLSRKREERTWGMODRLTYQTQOTRQTYVQPSFPLRLRIEENKSEA 1078  
 Qy 1262 RLAR-----TGETTIAQ-----QRASKVLLTSTDSQOS-----VILTSKONLS 1300  
 Db 1079 KAHRSRLAEVLQEDTGYTLPAELAVMLGLRLERLRQGSVEESQOMLQCGLTAEQGAQ 1138  
 Qy 1301 QEA-----YSAVGRKHSKANDASI-----LQYN 1323  
 Db 1139 LERGVYPERKLALYHCORGLPLGLISPRERETALYADEGNLSTSAQPLQOSLRFP 1198  
 Qy 1324 GERADPVGYTHLNGYRSYDPTLMRFHTPDSLSFPG-AGGINSYCLADPINSRPSG 1382  
 Db 1199 GQGVDESSGLYNNRNY--YDPLQGRYITQD--PIGLGGMNLYOYPL-NPIHIDPLG 1252

## RESULT 11

H91236  
 Rhsp core protein with extension [imported] - Escherichia coli (strain O157:H7, substra  
 C:Species: Escherichia coli  
 C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 09-Jul-2004  
 C/Accession: H91236  
 R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;  
 Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Rep. 8, 11-22, 2001  
 A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene  
 A/Reference number: A93629; M01D:21156231; PMID:11258796  
 A/Accession: H91236  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-1394 <HAY>  
 A/Cross-references: UNIPROT:Q8X385; UNIPARC:UP100000029D9; GB:BA000007; PTDN:BA838287.1;  
 A/Experimental source: strain O157:H7, substrain RMD 0509952  
 C/Genetic:  
 A/Gene: ECG4864  
 C/Superfamily: rhsp protein

Query Match 3.8%; Score 341; DB 2; Length 1394;  
 Best Local Similarity 21.5%; Pred. No. 1.9e-10;  
 Matches 260; Conservative 156; Mismatches 404; Indels 392; Gaps 63;  
 Qy 418 KTEWT-RSADNSYRSSEITETTFDESGNPLTKVI-----KDKTKQKLIIS--PS-----461  
 Db 218 RTQTFHREAAAGEBSGEITGYT-DGAGHFRVLVLTQAKABEAKQALSGSTESAPDPT 276  
 Qy 462 -----THMEYPPAGEVDNCPREP--YGFT-----RFRK 488  
 Db 277 LPQYTEXGRDNGTRLSAIVMLTHDEYP-----ENLPAAPLVRYGWTFRGELAVVYDSGK 331  
 Qy 489 KIITQTPDSEFK-----DDEKFIQYRYSLSIGSQSHVTLKIERHYSATQOLNSTL 539  
 Db 332 QVRSFTYDDKXRGVMVAHRHTGRPE--IRRYD--SDGRVTEQLNPAGLSYT-----379  
 Qy 540 FOYNTDSESLKRLKQEC--TKGENKTSYVHKFTYQDDTLQOQSHSITTHDNFTIH 597  
 Db 380 YQEKDRIITTDLSNREVLHTQCEGS-LKRVVAK-----EADGSVT 421  
 Qy 598 RQVRSRYTGRLPSTDTKDIVTQMSYDKLGRLLTRTNSGTPYANTLTVDYELANLQDD 657  
 Db 422 QSQFDA--VGRLAQDQDAARTEYSPDVVTGLTRIT--TPDGRASAFYY-----466  
 Qy 658 NRPPVYITTDVANGNQLRNEFDGAGRVASQCLKQSD-----GDGKYTHITQYDEQGRH 712  
 Db 469 NNSQLTSATGPGGLEIRREYDEGRGLIOETAPGDITTRRYNPHSDLCATEDATGSR 528  
 Qy 713 HTSTYSDYLTNGRQOQDPDKVHLSMSKSYNMGGIANTHMSYGVSEKITYTDP-TLWATK 771  
 Db 529 KMTWTSY--GQLSFTDSCGYVTRKDRFGGMVTVHBEESGLQRAYDSRQGLIAVK 585  
 Qy 772 QLOQSNMNVQTKGEVTVYTPSQOPIQLTLFDEAGHLQSCHTLTFRDGDWRVAKETDAIGC- 830  
 Db 586 DTQGHETRYE-----YNAAGDLTT--VIAPDG--SRNGTGYDAMGKA 623

Qy 831 -CTI-----YOYDNNRYIOITLPGDTIVNRKVAPESTDTLITDIRVNGISGOQTF- 881  
 Db 624 ICTTQGGITRSMERDAAGRVIRLSENGs-----HTFRYVDLDRLLQETGFDGRTORH 678  
 Qy 882 -DGLSRLTQSDGGRVWATYTSAGNDQCPSTVITPPDQFIHYOQPELDAVLQVANSNEI 940  
 Db 679 HDLTKLIREDEGLV-----THMHYD-EADLLTRHTVGET 714  
 Qy 941 TQGFSTNPTGAL-LKAVAGOSLPRIY--PSGL-LKMNINDMKMSIYM-----988  
 Db 715 AERWQYDE-RGMVLDLISHISEGRVLT-VHYGYDEKGLTGERQVTHHPQTEALMCHETR 772  
 Qy 989 ---TLNGEYTDLTGKISRDTHGRVTLQKDSIKTTLVNDLNRHIGSOVDTLAT 1045  
 Db 773 HAYNAQGLAN--RCITPDSLPVAVEMLYTG-----SGMLAGMKL 807  
 Qy 1046 GHMLTTTVEF--DGLNREIGRKLDSGHTLTDIQOSWTKTQOLANRIVKNGVLORTEQY 1103  
 Db 808 GD--TLPLVDFTRLRRLKRLRRRGVELTATYTPAQLOSHLNS-----LQYRDY 857  
 Qy 1104 SYDSRNLNOYKCDGACPTDKYGHSLVTONFTYDIYGNITACHTTPADGTEDHATKFA 1163  
 Db 858 TAMDNGELIR-----ISSFRQ-----TRSYSDSGRLTGVTTPAA-----NLDIRIP 900  
 Qy 1164 NPDPDC--QTEVHHTHPD-----MPDN-----IRKVDKXGRVINTID 1200  
 Db 901 YATDPAGNRLPD--PELHPDSTLSMPDPNRJARDAHYLYRDRGRRLTEKTLIPGVIRT 959  
 Qy 1201 NHGNTENFTYDTLGRLONGQ-----GSVGYDPLNLVSGQKTDLLDCEL-----1244  
 Db 960 DDERTRRYHNSQRLVHRYRTQYEBELVSRILYDPLGKRVAKRVARRRDLTGMSLS 1019  
 Qy 1245 -----YY-----RETMVNE-----VANGEMIRLR--TGETTIAQQA-----1276  
 Db 1020 RKPQVWYGMWDGRLFTIQNDRTRIQTITYQPSFTPLIRVETATGELAKTQBSLADALQ 1079  
 Qy 1277 -----SKULLTGDSQOSVILTS-----1284  
 Db 1080 QSGEDGGSVVPFPLVLOMLDRLESETLADRVSEBSRWLASGGLTVYQMOQOMDPVYTP 1139  
 Qy 1295 -----DKQNL-----SQEA-----YSAVGRKHSKSTAND--ASIIYNGERADPV 1330  
 Db 1140 AKKHLYHCHRGRLPLALISKGATKCAEYDWMNLANENHQLQLRLGQOYDEE 1199  
 Qy 1331 SGVTHLNGYRSYDPTLMRFHTPDSLSFPG-AGGINSYCLADPINSRPSGHLISQAW 1389  
 Db 1200 SGLYY--NRHRYDDPLQGRYITQD--PIGLKGMNLYTYPL-SPVNGMDPLGLYEFKSK 1253  
 Qy 1390 TGIGMGAGLIL 1401  
 Db 1254 NIDDIGIFALAM 1265

## RESULT 12

B65145  
 Rhsp protein precursor - Escherichia coli (strain K-12)  
 C:Species: Escherichia coli  
 C>Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 09-Jul-2004  
 C/Accession: B65145; S47701; B36902; A30092; I54935  
 R/Blatner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Col  
 A:Rose, D.J.; Mau, B.; Shao, Y.  
 A>Title: The complete genome sequence of Escherichia coli K-12.  
 A/Reference number: A64720; M01D:97426617; PMID:9278503  
 A/Accession: B65145  
 A/Status: nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-1411 <BLAT>  
 A/Cross-references: UNIPROT:P16917; UNIPARC:UP100001681C5; GB:AE000424; C  
 A/Experimental source: strain K-12, substrain MG1655  
 R/Plunkett, G.  
 submitted to the EMBL Data Library, March 1994

A:Reference number: 547666  
A:Accession: S47701  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1129, 'C', 1131-1411 <PLU>  
A:Cross-references: UNIPARC:UPI00001338AB; EMBL:U00039; NID:9466582; PIDN:AA18457.1; PIDN:AA18457.2  
A:Experimental source: Erikan K-12, substrain M6165  
R:Zhao, S.; Sandt, C.H.; Feulner, G.; Vlazny, D.A.; Gray, J.A.; Hill, C.W.  
J: Bacteriol. 175, 2799-2808, 1993  
A:Title: Rms elements of Escherichia coli K-12: complex composites of shared and unique  
A:Reference number: A36902; MUID:93259920; PMID:8387990  
A:Accession: B36902  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1397-1411 <ZHA>  
A:Cross-references: UNIPARC:UPI000017888C  
A:Note: sequence extracted from NCBI backbone (NCBIN:132073, NCBIP:132075)  
R:Sadosky, A.B.; Davidson, A.; Lin, R.J.; Hill, C.W.  
J: Bacteriol. 171, 636-642, 1989  
A:Title: the gene family of Escherichia coli K-12.  
A:Reference number: A91901; MUID:89123133; PMID:2644231  
A:Accession: A30092  
A:Molecule type: DNA  
A:Residues: 1-100 <SAD>  
A:Cross-references: UNIPARC:UPI000017888D  
R:Feulner, G.; Gray, J.A.; Kirschmann, J.A.; Lehner, A.F.; Sadosky, A.B.; Vlazny, D.A.; J: Bacteriol. 172, 446-456, 1990  
A:Title: Structure of the rhaA locus from Escherichia coli K-12 and comparison of rhaA w  
A:Reference number: I54935; MUID:90094253; PMID:2403547  
A:Accession: I54935  
A:Status: translated from GB/EMBL/DDBT  
A:Molecule type: DNA  
A:Residues: 1221-1411 <RES>  
A:Cross-references: UNIPARC:UPI000017888E; GB:M29717; NID:9147622; PID:9147623  
C:Comment: the rha core consist of two distinct parts: a large N-terminal core that is c  
C:Genetic8:  
A:Gene: rhaB  
A:Map position: 77 min  
C:Superfamily: rhaB protein  
C:Keywords: transmembrane protein  
F:1-36/Domain: signal sequence #status predicted <SIG>  
F:127-141/Product: rhaB protein #status predicted <MAT>  
F:128-55/Domain: transmembrane #status predicted <TM>

[illegible]





C/Genetics:  
A/Gene: rhea  
C/Superfamily: rhea protein  
C/KeyWords: transmembrane protein  
F/1-26/Domain: signal sequence #status predicted <SIG>  
F/23-1377/Product: rhad protein #status predicted <MAT>  
F/28-55/Domain: transmembrane #status predicted <TM>

Query Match 3.7%; Score 331.5; DB 2; Length 1377;  
Best Local Similarity 20.8%; Pred. No. 6.2e-10;  
Matches 288; Conservative 170; Mismatches 450; Indels 477; Gaps 75;

Qy 211 LFPQKEGY-RTELRFLNRLQNSIHNSLGNENPLTWSFGYTPGKNGILGQWITSMTAP 269  
Db 126 LFPGE-DGYSRSESLMVR--GSVAKDEGRLAALNQ----- 160  
Qy 270 GGLKEVNYG-----NNNG-----HHFQSANL---PVLPLYTLMKQVPGAGQ 310  
Db 161 -ALPEELRLSPHRYLATNSPQSPWMLGWCERPEADEVLPAFLPPRYVLGLVDRFGRT 219  
Qy 311 PALQAEYSYSHNVGSGNSGIMNNKDLNGLMTENVGSTRYKDKGHDQIVRIE 370  
Db 220 QTFHREKAGFSGEITGVTDGAMH-----FRLVL-----TTOQR----- 255  
Qy 371 RYNNHMLTSECKQNGYIQTETAYVAIIGHNPDSPQSOFOPLKT---KTETWRGADN 427  
Db 256 -----AEKAAQ-----AISG---GTBPFAF---PDLPGTYEGR--DN 287  
Qy 428 SYNSEITETTPDESQNLPLVYIKDKTKQKISPTHWYPPAGEVDCPEP---YGFT 484  
Db 288 GIRLSAVMLTHDPE-----YP-----ENLPAALVRYGWT 317  
Qy 485 -----RKYKIIQTPYDSEK-----DDEPKIORYSLISQSHVTLKIE 525  
Db 318 PRGELAVVYRSGKQVRSFTYDDCKRGMVAHRRYGRPE--IRRYD---SDGRVTEQLN 372  
Qy 526 ERHYSATQLNSTLFOYNTDKSELGRLLKQTEC--TGKNGKTVSVVHKFTYTKQDDTLQ 583  
Db 373 PAGLSY-----YQYKDRITITDLSLRREVHTGEGAG-LKRVYK----- 413  
Qy 584 QSHSITTHDNFTIHRSGVRSRYTGRFSDTDTKDVIYQMSYDKLGRLLTTLNSGTPYAN 643  
Db 414 -----EHADGSVYQSQFDA--VGRLEAQTDAGRTESPDVVTGLITRLT--TPGR 462  
Qy 644 TLTVDYELNQLQDNRPPVITTTDVANGQLNRPDAGRHVSQCLQSD-----GDGF 698  
Db 463 ASAFYYVNHQO-----LTSATGPDGLRLREYDELGRLLIETAPDGDITRRYRDNPH 514  
Qy 699 YTHITQOYDEQGRHHTSYSDYLTNGRQOTPDKVLHLSMSKSYDNMGQIANTHWSYGVSE 758  
Db 515 SDLPKATEDATGSKRTWRSY--GQLLSFTDCSGYTRYDHDHFGOMTAHVHREGLSQ 571  
Qy 759 KITVDPI-TLTATKQLOSNNSNVQTKAEVTTYTPSQOPIQITLLEDEAGHLQ-----SC 810  
Db 572 YRAVDSRGLIADVDTQGHETRYE-----YNAGDILAVIAPDGSR 612  
Qy 811 HTLTRGMDVRRKETDAIGCCTIYQDNVNRVIOITLPDGTIVNRKTAAPSTDTLITDIR 870  
Db 613 NGTOYAMGAVALRTQG--GLTRSMEDYDAAGVIRLITSENGS---HTTFRYVDLDRLLQ 666  
Qy 871 VNGISLQOQTF--DGLSLRTQSDGGRVMAVYASAGNDQCPSTVITPDGQFIHQYOPEL 928  
Db 667 ETGPDGRTQRYHNDLTGKLRSEDEGLV-----THWHTD-EA 702  
Qy 929 DDAVLQVANSNEITQGFSPNPTVGL--LKAVALGOSLTPYY--PSGR--KMNINDM 981  
Db 703 DRLTRFTVKGETAERWQYDE-RGWLTDISHISEGRVA-VHYRYDEKGRLTGERQVAVHP 760  
Qy 982 KKMSTLM-----TLRGLNGYTDLTGTLTKISDPTGRVTOIKDSIKTTLNVDL 1033  
Db 761 QTEELLMQHETRAHYNAQGLAN--RCIPDSLPAVEMLTYG-----SGYLAGKLGAD-- 809  
Qy 1034 RHIGSQVTDLATGMLTTVEF--DGLNREIGRLCDSSGHTLIDQOSWLKTQOLANRIV 1091

Db 810 -----TPLYEYTRDRRLHRETL-----SGRGEYLTATYTPAGLOSQ-- 846  
Qy 1092 KLVGLQRTQEQYSDSRNRLNQKCGACPTDYGHSIVYQNTFYDIYNTACHTTFA 1151  
Db 847 HNLNLS-DRDYTWNDNGELIR-----ISSPRQ-----TRSYSTGRLTGVHTTAA 893  
Qy 1152 DGTEDHATPKFANPTDPC--QLTEVHHTHPD-----MPDN-----IRLKYDAKGVIN 1197  
Db 894 -----NIDIRIPYATDPAQNRLLPD-PELHPDSTLSMPDNRILADAHYLYRYDRHGLTE 947  
Qy 1198 ITD-----NHGNTENFTYDLGRLQNGQ-----GSVYGYDPLNRLVSQKTDI 1239  
Db 948 KTDLIPRGVARTDDETRHRYHDSQHRLVHTYTYQYEPVLESRYLVDLGRVAKVWR 1007  
Qy 1240 LDCEL-----YF-----RETMVNE-----VRNGEMIRLIR--TGFTI 1270  
Db 1008 RERDLTGMSLSRKPQVLTWYGMDDRLTTIQNDRTRIQTIIYQSGSPFLIRVETATGELA 1067  
Qy 1271 IAOORA-----SKVLLGTDSQSVL-----TS 1294  
Db 1068 KTORSLADALQSGGEGGSVFPVLYQMLDLRLSEILADRVSESRRLASGGLTVE 1127  
Qy 1295 DKON-----LSQEA-----YSAYGKHSTAND--AS 1318  
Db 1128 QMONQNDPVYTPARKIHLHYCHDRGLPLALISKEGTEGCAEYDEWGNLNEBNPHOLOQ 1187  
Qy 1319 ILGNGERADPVSGVTHLNGYRSYDPTLMRFHTPDSLSPFG-AGGINPYSYCLGDPINR 1377  
Db 1188 LIRLPQGOYDEBSGLYV--NRHRYVDPLOGRYITQD--PIGLKGGWNPYQYPL-NBVTN 1241  
Qy 1378 SDPSG 1382  
Db 1242 TDPG 1246

Search completed: January 30, 2006, 09:49:50  
Job time: 40.1754 secs

**THIS PAGE LEFT BLANK**

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2006, 09:40:48 ; Search time 133.373 Seconds  
(without alignments)  
8849.971 Million cell updates/sec

Title: US-09-889-874A-23

Perfect score: 8879  
Sequence: 1 VYIKFLKLFRRITMSDNNF.....PRKILGRTEKTVKPTFRP 1673

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues  
Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt 05.80.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8815	99.3	1660	09EVR7_XENBV	09EVR7 xenobacillus
2	1632	18.4	1625	07MB38_PHOIL	07mb38 photobacillus
3	1615	18.2	1578	04ZUT5_PSEBY	04zut5 pseudomonas
4	1614	18.2	1562	0883V9_PSEBY	0883v9 pseudomonas
5	1598	18.0	1385	07NMW2_CHRYO	07nmw2 chromobacter
6	1583	17.8	1775	07NA47_PHOIL	07na47 photobacillus
7	1553	17.5	773	09S6J1_COXBU	09s6j1 coxiella bu
8	1530.5	17.2	774	052880_COXBU	052880 coxiella bu
9	1490.5	16.8	1682	04ZUT9_PSEBY	04zut9 pseudomonas
10	1416	15.9	1632	088LP7_PSEBY	088lp7 pseudomonas
11	1214	13.7	709	045948_COXBU	045948 coxiella bu
12	1037	11.7	1290	088LP2_PSEBY	088lp2 pseudomonas
13	932.5	10.5	982	0883W6_PSEBY	0883w6 pseudomonas
14	920.5	10.4	1669	087VG6_PSEBY	087vg6 pseudomonas
15	919.5	10.4	505	045905_COXBU	045905 coxiella bu
16	919.5	10.4	526	09X6Z6_COXBU	09x6z6 coxiella bu
17	916.5	10.3	526	052883_COXBU	052883 coxiella bu
18	912.5	10.3	528	045949_COXBU	045949 coxiella bu
19	884	10.0	589	045902_COXBU	045902 coxiella bu
20	648	7.3	762	0931F8_3VIRB	0931f8 vibrio pom
21	552	6.2	370	07NVR8_CHRYO	07nvr8 chromobacter
22	486	5.5	389	07NVA0_CHRYO	07nva0 chromobacter
23	478.5	5.4	310	045904_COXBU	045904 coxiella bu
24	450.5	5.1	2217	08TP72_METAC	08tp72 methanocarc
25	450	5.1	380	07NM05_CHRYO	07nm05 chromobacter
26	439	4.9	2554	072OR5_LEPTIC	072or5 leptospira
27	432	4.9	2334	WAPA_BACSU	007833 bacillus su
28	429.5	4.8	336	045946_COXBU	045946 coxiella bu
29	420.5	4.7	2364	082R58_STRAW	082r58 streptomyces
30	418.5	4.7	1835	06MD86_PARUV	06md86 parachlamydia
31	414.5	4.7	2221	073B23_BACCI	073b23 bacillus ce

32	405.5	4.6	2221	04MM58_BACCE	04mm58 bacillus ce
33	403	4.5	371	04ZK7_PSEBY	04zk7 pseudomonas
34	396	4.5	1513	07NY44_CHRYO	07ny44 chromobacter
35	395	4.4	820	082OR1_STRAW	082or1 streptomyces
36	395	4.4	2246	073C66_BACCI	073c66 bacillus ce
37	393.5	4.4	1976	08DAR5_VIBU	08dar5 vibrio vuln
38	393	4.4	363	04ZLS4_PSEBY	04zls4 pseudomonas
39	392	4.4	1250	082RX1_STRAW	082rx1 streptomyces
40	391.5	4.4	2370	082RE3_STRAW	082re3 streptomyces
41	391	4.4	331	04ZK8_PSEBY	04zk8 pseudomonas
42	388.5	4.4	1826	0987Z7_RHIO	0987z7 rhizobium 1
43	386	4.3	2167	092BK5_LISTIN	092bk5 listeria in
44	386	4.3	2224	081U00_BACAN	081u00 bacillus an
45	382.5	4.3	1117	08CK70_STRCO	08ck70 streptomyces

## ALIGNMENTS

RESULT 1					
ID	09EVR7_XENBV	PRELIMINARY;	PRT;	1660 AA.	
AC	09EVR7;				
DT	01-MAR-2001 (TrEMBLrel. 16, Created)				
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)				
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)				
DE	Nematocidal protein 2.				
GN	Name=xnp2;				
OS	Xenobacillus bovienii.				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;				
OC	Enterobacteriaceae; Xenobacillus.				
OX	NCBI_TaxID=40576;				
RN	[1]				
RP	NCU02107.2 SEQUENCE.				
RC	STRAIN=173;				
RA	Morgan J.A.W., Quesley M., Ellis D., Jarrett P.;				
RL	Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.				
DR	EMBL: AJ296651; CA19493.1; -; Genomic_DNA.				
DR	InterPro: IPR006530; YD.				
DR	Pfam: PF05593; Rhs_repeat; 5.				
DR	TIGRFAMs: TIGR01643; YD repeat 2x; 4.				
SQ	SEQUENCE 1660 AA; 185738 MW; AFA8EA20AD70B164 CRC64;				
Query Match 99.3%; Score 8815; DB 2; Length 1660;					
Best Local Similarity 100.0%; Pred. No. 0;					
Matches 1660; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	14 MSDNNEFTQANNFTSAVSGVDPRGLVNIQITLGHIVGNGLGPTLPYLLSYSPINKT				73
DB	1 MSDNNEFTQANNFTSAVSGVDPRGLVNIQITLGHIVGNGLGPTLPYLLSYSPINKT				60
QY	74 DIFRGIGFNGFGLSVYDRKNSLSLSTGENYKVIETDKTVLQOKLDNLPKDKLNCY				133
DB	61 DIFRGIGFNGFGLSVYDRKNSLSLSTGENYKVIETDKTVLQOKLDNLPKDKLNCY				120
QY	134 RIHKSGDIEVLGPNNAADLVKPKLLNPAGHAIYIDNFPATOPRLRIYDDLDGHD				193
DB	121 RIHKSGDIEVLGPNNAADLVKPKLLNPAGHAIYIDNFPATOPRLRIYDDLDGHD				180
QY	194 IPLLNTLEYGLIKITLITLPFGQKRGYRTELRLNRLNLSIHNSLGNENLWTFGYTPI				253
DB	181 IPLLNTLEYGLIKITLITLPFGQKRGYRTELRLNRLNLSIHNSLGNENLWTFGYTPI				240
QY	254 GKNGIIGQWITSTWAPGLKELTVNYSNNQHHFPOSANIPVLPYVTLKQVPGAGPAI				313
DB	241 GKNGIIGQWITSTWAPGLKELTVNYSNNQHHFPOSANIPVLPYVTLKQVPGAGPAI				300
QY	314 QAEYSTSHYVGGSGNGINNNKLDNLYGIMTEYVNGSTSRBYKXEGHDOIVRLERTY				373
DB	301 QAEYSTSHYVGGSGNGINNNKLDNLYGIMTEYVNGSTSRBYKXEGHDOIVRLERTY				360
QY	374 NNYHLTSECKQONGYIQOTETAYVAILIGHNFDSPQFOLPPTKTTETWRSANNSYSEI				433

Db 361 NNHLLTSECKQNGYIQTETETAYAILGNFDSQPSQPOLPKTETETWBSADNSYRSEI 420  
Qy 434 TETTFDSSGNPLTVIYIDKDKTKIISPTHEWEYPAPEGVNCPPEYGFTRPFKXIIOT 493  
Db 421 TETTFDSSGNPLTVIYIDKDKTKIISPTHEWEYPAPEGVNCPPEYGFTRPFKXIIOT 480  
Qy 494 PYDSEFPKDDPEKFIQYRYSLSIGSOSHWLTKIEERHSATQOLNLTLPQYNTDKSELGRLL 553  
Db 481 PYDSEFPKDDPEKFIQYRYSLSIGSOSHWLTKIEERHSATQOLNLTLPQYNTDKSELGRLL 540  
Qy 554 KQTECTYGENGKITYSVVHKFTYTKODDTLOQSHSITTHDNFTIRSOVRKATYRLASDT 613  
Db 541 KQTECTYGENGKITYSVVHKFTYTKODDTLOQSHSITTHDNFTIRSOVRKATYRLASDT 600  
Qy 614 DTQOIVTQMSYDKLGRLLTTRTNGSTPYANLTVYDEYLNLODNRPPEVPTTTDVANGQ 673  
Db 601 DTQOIVTQMSYDKLGRLLTTRTNGSTPYANLTVYDEYLNLODNRPPEVPTTTDVANGQ 660  
Qy 674 LRNEFDGAGRHSOCLKDSGDGKFTYTHIQOYDEQGRHHTSTYSDYLTNGROQTDPDKV 733  
Db 661 LRNEFDGAGRHSOCLKDSGDGKFTYTHIQOYDEQGRHHTSTYSDYLTNGROQTDPDKV 720  
Qy 734 HLSMSKYDNMGQIANTHWSYGVSEKTVDPITLPAIKOLOSNNSNNVQTKVETTYPSQ 793  
Db 721 HLSMSKYDNMGQIANTHWSYGVSEKTVDPITLPAIKOLOSNNSNNVQTKVETTYPSQ 780  
Qy 794 QPQITTLFDEAGHOSCHTLTRDMDRVRKETDAIGCTIYOYNNRVIOITLPDGTIV 853  
Db 781 QPQITTLFDEAGHOSCHTLTRDMDRVRKETDAIGCTIYOYNNRVIOITLPDGTIV 840  
Qy 854 NRKTAPESTDTLITDIRVNGISLSGOQTFDGLSRLOSODGRVWATYYSAGNDQCPSTVI 913  
Db 841 NRKTAPESTDTLITDIRVNGISLSGOQTFDGLSRLOSODGRVWATYYSAGNDQCPSTVI 900  
Qy 914 TPDCQFIHYQYQPELIDAVLOVANSNETQOQSYNPVTGALLKAVAEQSLTPYIPSGRL 973  
Db 901 TPDCQFIHYQYQPELIDAVLOVANSNETQOQSYNPVTGALLKAVAEQSLTPYIPSGRL 960  
Qy 974 KMENINIMKMSYMTLRGLNGVYTLNGTIOKISRDTGSRVTOIKOSIKTTLYNDYDLN 1033  
Db 961 KMENINIMKMSYMTLRGLNGVYTLNGTIOKISRDTGSRVTOIKOSIKTTLYNDYDLN 1020  
Qy 1034 RHISQOVTDLATGHLATTVVEFDGINREIGRKLCDSSGHTLIDIOQSWLTKOOLANRYKL 1093  
Db 1021 RHISQOVTDLATGHLATTVVEFDGINREIGRKLCDSSGHTLIDIOQSWLTKOOLANRYKL 1080  
Qy 1094 NGVLOQTEQSYSDSRNRLNOYKCDGACPTDKYGHSHIVTONFTYDIYGNITACHTTFADG 1153  
Db 1081 NGVLOQTEQSYSDSRNRLNOYKCDGACPTDKYGHSHIVTONFTYDIYGNITACHTTFADG 1140  
Qy 1154 TEDHATPEFANPTPCQULEVHHTHPMDNIRLKYDAGRVINTDNHGNTENFTYDTL 1213  
Db 1141 TEDHATPEFANPTPCQULEVHHTHPMDNIRLKYDAGRVINTDNHGNTENFTYDTL 1200  
Qy 1214 GRLQNGQSYGYDPELNLVSOQTDITLDCELYREETMLVNEVRNGEMIRLRTGTTIIAQ 1273  
Db 1201 GRLQNGQSYGYDPELNLVSOQTDITLDCELYREETMLVNEVRNGEMIRLRTGTTIIAQ 1260  
Qy 1274 QRAKAVLLTGTSQGSVILTSQKONLSQAYSAVYKHSSTANDASILLGYNGERADPYSGV 1333  
Db 1261 QRAKAVLLTGTSQGSVILTSQKONLSQAYSAVYKHSSTANDASILLGYNGERADPYSGV 1320  
Qy 1334 THLNGVSYDPTLMRFHTPDSLSFPGAGINPYCYCQDPINSDPBGHLSQWQMTIG 1393  
Db 1321 THLNGVSYDPTLMRFHTPDSLSFPGAGINPYCYCQDPINSDPBGHLSQWQMTIG 1380  
Qy 1394 MGIAGLTLITATGMAIAAGGIAAIAASTSTTALAFALSVTSITITSIVGALADAPK 1453  
Db 1381 MGIAGLTLITATGMAIAAGGIAAIAASTSTTALAFALSVTSITITSIVGALADAPK 1440  
Qy 1454 ASSIIGVWSMGGAAGLAESAIGKGTKLATHLGAFAEDGENALLKSTSESSRIKMGVTRS 1513  
Db 1441 ASSIIGVWSMGGAAGLAESAIGKGTKLATHLGAFAEDGENALLKSTSESSRIKMGVTRS 1500

Qy 1514 LDREIVANBERGOYIKDHSRGYTDNFMKGEOALIVHGDKDGFLYHTEGNKNGKPYTRH 1573  
Db 1501 LDREIVANBERGOYIKDHSRGYTDNFMKGEOALIVHGDKDGFLYHTEGNKNGKPYTRH 1560  
Qy 1574 TPQOLVDYLKONNIVDLTQGGDKPVHLLSCYKSSGAAADMAKYNRPVLAYSKPTISQ 1633  
Db 1561 TPQOLVDYLKONNIVDLTQGGDKPVHLLSCYKSSGAAADMAKYNRPVLAYSKPTISQ 1620  
Qy 1634 GLARIEKDFELKSTYASVDPKRIILGRTEKYAPKTFRP 1673  
Db 1621 GLARIEKDFELKSTYASVDPKRIILGRTEKYAPKTFRP 1660

RESULT 2  
Q7MB38\_PHOHL PRELIMINARY; PRT; 1625 AA.  
ID Q7MB38;  
AC Q7MB38;  
DT 01-MAR-2004 (TREMBlrel. 26, Created)  
DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)  
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
DE Similar to the nematocidal protein 2. Probable membrane protein.  
GN OrderedLocustNames=Dlu2222;  
OS Photobacterium luminescens (subsp. laumondii).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Photobacteriaceae;  
OX NCBI\_TaxID=141679;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=TT01;  
RX MEDLINE=22957627; PubMed=14528314; DOI=10.1038/nbt886;  
RA Duchaud E., Ruenick C., Frangeul L., Buchrieser C., Givaudan A.,  
RA Taourit S., Bocs S., Bouraux-Eude C., Chandler M., Charles J.-F.,  
RA Daes E., Derose R., Derzelle S., Freydisse G., Gaudriault S.,  
RA Medigue C., Lanois A., Powell K., Siguler P., Vincent R., Wingate V.,  
RA Zouine M., Glaeser P., Boemare N., Danchin A., Kunst F.,  
RA "The genome sequence of the entomopathogenic bacterium Photobacterium  
RT luminescens.";  
RL Nat. Biotechnol. 21:1307-1313 (2003).  
DR EMBL, BX571866; CA514515.1; -; Genomic\_DNA.  
DR Photobacter; pluz2222; -.  
DR InterPro; IPR006530; YD.  
DR Pfam; PF05593; RBS\_repeat; 7.  
DR TIGRFAMs; TIGR01643; YD\_repeat\_2x; 3.  
KW Complete proteome.  
SQ SEQUENCE 1625 AA; 180118 MW; BFA362F4BCB3BA97 CRC64;

Query Match 18.4%; Score 1632; DB 2; Length 1625;  
Best Local Similarity 31.0%; Pred. No. 1,48-79;  
Matches 494; Conservative 245; Mismatches 653; Indels 204; Gaps 59;

Qy 18 NEFTQANNFTSANSGGVDPRTGLYNTQITLGHVGN--GNLGPFLPLTSLYSPLNTKDI 75  
Db 4 NDIYSNAPNFQSYINTVDPRTQYSANINITLRPNVNGIYGV--LNLFSFPLTILN 61  
Qy 76 GFGIGFNGSLVYDRKNSLSLSTGENYK---VIEETKTVLQOQKLDNLFERDLKENC 132  
Db 62 GFGIGFNGSLVYDRKNSLSLSTGENYK---VIEETKTVLQOQKLDNLFERDLKENC 120  
Qy 133 YRIHKSGLIEVLTFGNNAFEDLKVPKLLNPAGHAIVIDWNEPATOPRLNRIYDLDGH 192  
Db 121 FYVYNKQGIIEFLTKRIGSS--DIKTVALEPDS-----EVEDLIYNS 161  
Qy 193 DIPLNLEYQGLKTIITLT--FPQKGEKRYTELRL--NRQLSINHPFSGNENPLTWSGY 250  
Db 162 RFALSEIKRYMTGTYIKLVNNGNN--CTSVETPDNNISAKLAFDYRNDLITVTPY 218  
Qy 251 TPICK-----NGILGOWITSMAPGLKETVYNSNNQGHFPQSANLPVLP 297  
Db 219 DASGPIDSAKPKMTYQTLKIF--PVISAFFPTGYVELVSVKEN--GKAVTDIESIPYAA 275  
Qy 298 YVTLMKQVPGAQPAIQAEYSYTS--HNYVGGSNGI--MNNPLDNLXGLMTEYNTGSTESR 355  
Db 275 YVTLMKQVPGAQPAIQAEYSYTS--HNYVGGSNGI--MNNPLDNLXGLMTEYNTGSTESR 355



```

Db 276 ALTTC--PAGCPAISKSEYSSVHNFLGYSGRTSFDSSQDNLVLTGKTYSSIE-- 330
Qy 356 RYDKGCHDQIVIERTYNNHLLTSECKQONGIQTETAYVAIGHNDSQSQQLP 415
Db 331 --RYLNGQNYISYERVFDFKHLMTKAKQDNKRITTEIYNEDPKSPSEOPENQOP 388
Qy 416 K-----TKTEWMSAUNYRSEITETTFDESGNPLTKVIKDKTKOKIISPTWMEYYP 469
Db 389 SHVLTRTDLOT-----NTSRESVNKSDWMGNTL-ITETSGIQK-----EYVYFV 436
Qy 470 AGEVNCPPPEYGFTRFKKIIQTPYDEFPDDEKEIYRYSLIGSQSHVLTKEIRHY 529
Db 437 NGEKNCPADPLGFSRLKSVTGKSPDAQSVANRVTSTYQQLPFTGAYVK--EYVS 494
Qy 530 SATQOLNSTL---FOY---NTDKSELGRLLKQECTGKNGKTYSVHAKTKTYQODTLQ 583
Db 495 KASBETIDSKIVRFNPNVNSPTNKS--GSLAKITSVMNQ-----QYTTKREYSDSEMT 548
Qy 584 QSHSITTHDNFTIHRQVRSRYTGRLPSPDPTKDIYOMSYDKGRLLTTLNAGTPYAN 643
Db 549 TNSVTGFPDGTWMSKNVTSIYTRKQARKVDVNHVITDOSTYLSGRITIGQIIDGTYKEI 608
Qy 644 TLTYVELANLQODNRPPVITTTDVANGQLRNEPDAGRHVSQLKDSQ-----D 695
Db 609 KRSYIYQPGDENDFWP-VWIEIDSGIRKTHYDGMGRICIEQDDDGWGTSGIYQ 667
Qy 696 GKRYTTHQOYDEGGRHHTSYSLTN-----GRQOTDPKHLNLSKSYDNWQIANT 750
Db 668 GYRKVYLAQYDVYLGQLVKSIISNDWLMSANPLRLTTP--LVTTKYQYDGMGNVYST 725
Qy 751 HMSYGVSEKLTVPDITLTATK-----QLOSNNNVQTKGVNTYTPSOPIOTLTDEA 804
Db 726 EYDGRLELEIHDTITTIQGVKGLMNLQNNF-----EGPASKIVYPPD 773
Qy 805 GHLQSCHTLRDGMWRKSTDAIGQCTIYQDYNVNEIQTLPDGTIVNRKXAFESTD 864
Db 774 GAIYSTRYHYDGRVTEITDAGVATQIEYDFRIVKKTLPDRITLSAVASFHEE 833
Qy 865 LIIDIRVNGISLGQOTDGLSRLTQSDGGRWAVYTSAGNDQCPSTVITP--DQGITIH 922
Db 834 LISALVNGQLSLVYDGLGRVTRDYGKRTYELGSGQDK-PISQVTPAHNKKONIDY 892
Qy 923 QYQBELDAVLQVANSNEITQOFSYNPVTGALLKKAIVAGOSLTP--YPSGRLLKMNEN-D 980
Db 893 LY--ALQSVMSKFTTETSQONFSYQKTGALLS--TEGVOSQNSYSPSGVLLQIESSRD 949
Qy 981 MKQMS---YLMTLRLGNGYTLDTGTIOKISRDTGHRVTOIKDSIKTTLVYDNLNHIQ 1037
Db 950 NKPISSGDYRYTMSGLIQRHKDSFAHDHVSYDAEGLVTEGSSQYATFEYDVGRLIT 1009
Qy 1038 SOVTLATGMLTTTVEFDGLNREIGKLCDSGHTLIDIOQSWLKTQOLANRYKLVGL 1097
Db 1010 TTYKDTTSLQALTKIEYDVFDRERIKRSLISDFSIOV-ITLSYKNNQISQRLISIDGV 1068
Qy 1098 ORTEOVSYSRNLNOYKCDGAECPDCKYGHSTVTONFTYDIGNFACHTTPADGTEH 1157
Db 1069 MKERIVYDNORLSOYCEBQSPVDHTRVLSQOYHNDQWENIRLONTRADGET- 1127
Qy 1158 ATRKFNPTDPCQLEVHTHPMPDNIRLKYDKAGHVINIT-DNHGNTENFTYDLGRL 1216
Db 1128 VDHFSQ-ADPTQIRI---TSDKOIELSYDANG--NLTRBEKQOT--LIIDQNNRL 1177
Qy 1217 ---ONGQASV---YGYPLNRLVSO--KTDTLQCELYYRRTMLVNEVRANGEMRL-- 1264
Db 1178 VQVYKDSKGNLVCOYQDALNKLTAQVLANGVNNO--YYASGVANAVOLGEBETITWSSDK 1236
Qy 1265 -----RTGETTIAQGRASKVLTLGTDSQSYTLTSDKONLSQEYASVYGHKXSTAND 1316
Db 1237 QRLGHQSTKQGESYTYQ-----GTDHNSYTIASQENELMALSYTYGFRSLI--- 1285
Qy 1317 ASILGVNGEADPVSVTHLNGRYSDPTLMRPHTPDLSLPGAGGINPYCYLADPIN 1376
Db 1286 SSIPGLNGAQVDVPTGVYFLNGYRVFNPVLMRHFSPDSWSPFGGGINPYCYLADPIN 1345

```

```

Qy 1377 RSDPSGHLISWQAMTGIQMGITAGLILTIAT--GGMAIA-----AAGTAAATSTTALAF 1430
Db 1346 RIDINGHLISAGLIGLIVLAGIIVGVSLGAPALISAGIILAGALGALASTAFVTA 1405
Qy 1431 GALSVDTSIYSVSGALDEASPRASSIIGVNSMGMAAGLAEASAI-----KGGTK 1480
Db 1406 TVIGLAADSIGIASMALSEDPKTAGILNINISGLVLSRGISAIPTTSSLIKARSQSQ 1465
Qy 1481 LATH--LQAF-AEDGENALIKSTESSRIKRWYTRS 1513
Db 1466 VASTSVIGSVPIEFGEIA-----SRSR-RMDIALS 1495

RESULT 3
Q4ZUT5_PSESY
ID Q4ZUT5_PSESY PRELIMINARY; PRT; 1578 AA.
AC Q4ZUT5;
DT 13-SEP-2005 (T-EMBLrel. 31, Created)
DT 13-SEP-2005 (T-EMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (T-EMBLrel. 31, Last annotation update)
DE YD repeat.
GN ORFNames=Psy2044;
OS Pseudomonas syringae pv. syringae B728a.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=205918;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=B728a;
RA Loper J.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=B728a;
RA Feil H., Feil W.S., Lindow S.B.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; CP000075; AAY37087.1; -; Genomic_DNA.
DR InterPro; IPR006530; YD.
DR Pfam; PF05593; Rhs_repeat_6.
DR TIGRFAMs; TIGR01645; YD_repeat_2x; 2.
SQ SEQUENCE 1578 AA; 177054 MW; 51A4B576ABCC10C9 CRC64;

Query Match 18.2%; Score 1615; DB 2; Length 1578;
Best Local Similarity 30.4%; Pred. No. 1,1e-78;
Matches 481; Conservative 251; Mismatches 680; Indels 168; Gaps 48;

Qy 14 MSDNNEFTQANNFTSAVSGVDPRGTGYNIOITLGHIVGNLIGPTLPLTSSYPIAKT 73
Db 1 MTAISTVHSAFNFLSYQGVDPRTGQYVAITLTPVKNTGRLRGPMPLALNNPILNRQ 60
Qy 74 DIGFGTGFNGSLSVYRKNSLSLSTGENYKVLET-DKYVLAQOKLDNLRFEDLKEHC 132
Db 61 DSGFGGLWNLQLSQYDPGNQIVLSLSEGTFFKVGSTSDQMLSEKIDSFHLKYE-DETH 119
Qy 133 YRIHKSGLIEVLTGFNNNAFDLKVPRKILNPAIGHAIYIMNBEATQPLRIATYDDLDG 192
Db 120 FRVMSHSGVLVLELHAGSGANRYALPVRYAPRPGHSTLDYAFSGYQMLSEVTTDD-SGQ 178
Qy 193 DIPLANLEYQGLIKTI-LTLPQQKEGYRTFLPLNRQLNSINFSIGENPLTWSFGYT 251
Db 179 --LTLGISREDTAVLTAMTTGEGPDATFKVVLGSGSHRVRI---ELPTDNAASWRFVYS 233

```

```
QY 252 PIGKNGILGOWITMTAPGGILKETVNTYNNNOGHFPOSANL.PVL.PYVTLTKQVPGAGOP 311
D 234 LINEH-----LCMTSVETPVGREGIEFTYTDG--GHQFPAARAPRLPVTHRTTRTTPGIGQA 287
QY 312 AIOAEYSY-----TSHNYVGGSGNGIW--NNKLDNL.YGLMTE--YNGSTESRRYKDEGH 363
D 288 QMDVFSYQOSAGNERHFLAGLDIAMEDNGLDNLRYLAGELEYESTETLRLQVAGKP 347
QY 364 DQVRIERTNNHLLTSECKOQNGYIQTETAYALIGNFPDSQBPQOLPKTKETWR 423
D 348 DAERSIRRTNQFLLTRETTRKMLTVMQVTRYYIYPGQFPYQAPYCOLPEXQVTTMR 407
QY 424 -SADNSY-RSEITTFDESIGNPLTKYIKDKTKOKIISPSTHMEYYPAGEVNDCEPEPY 481
D 408 LSSGSVPRSEMVSTYDKGNLLTQT-----QANGMEVVS---EWMYAAAGE-DOCPDPPE 459
QY 482 GTFTRFKKIIQTPYSEFKDPEKFIQYRYSLIGSQSHVTLKIEERHYSATOLNS--- 537
D 460 GFVFMKSSKIIIPASDYGNAPV.LITR.YRYALPAVGGQLQOQWILAPQSETLMLQABEDE 519
QY 538 -----TLFQYNTDSEL---GRLLKQTECTGNGKT-----YSVNHFTYTKODDTLQOS 585
D 520 YEBORLTYYEFDEBNMALLHGR-LKQRVINGENATSTEVASTPDESEYK--TVLQTV 576
QY 586 HSIITTHDNFTIHRQVRSRYTGR.LFSPDTKDIYQMSYDKLGRLLTRTLNSGTP-YANT 644
D 577 QTVTFGDAKAKVTLLEBSLNGEBPLARDNNVIRYVBSLRLVLSBTVAPNDIYQAT 636
QY 645 LTYDYELNLIQDDNRPPFVITTTTVNGNQLNEFDGAGRHSQCLKSDGDG--KYTYI 701
D 637 RSEYEQ.CAKATDQATQ---TLIVKGAVKCSHF.DGLNRV.IHERDADKPARVYDKRQ 693
QY 702 HTQOYDEGRHHTSYSDYLINGBOQDIPDKVHLMSKSYNMKQIANTHMSYVSEKIT 761
D 694 YAAVYDAMGLVNEFEFDMW---GOODLALKSVY-----EYDWEBOQRCVTGADNIKTFEE 746
QY 762 VDP1-TLATKQLOSNNVQTKREVITYTPS-----QQPIQTLFPEAGHLQSCHTLTLR 815
D 747 TDP1GTASQOPVQBSWIEGTDGKGSFTETWNLFEPQFRSRKODAGTSVLSQVHY 806
QY 816 DGMWRVREKETA1---GQCTIYQYDNVNRV.IQTL.PDGTIYNRKXAPSTDTLITDIRV- 871
D 807 DGLGRVREIIVGVNNGGRRTTYVDVDFRVEIESTLPSAIVRSASVSHSDELDLISIGIE 866
QY 872 -NG--ISLQOCTFPGLSRLTOSQOGGRWATYASGNDQCSTYITTPGQGIHYOYQEL 928
D 867 HNGKSIVLGEQSPGLDRLVSSVYTGGRKRLTVYDVGLIQ-PRYVTLPEKRIKIDYEYKQV 925
QY 929 DD-----AVLQVANSNEITQOFSYNPYTGALLKAVA 958
D 926 GDEPMRVRVQSEABEPDTRREVDP.LKQSGPGAAPAR.TAPGVAADYTYDKQNALHSEE 985
QY 959 EGQSLTPIYPSGRKLMENINDMKK--MSYIMTLRGLENGYYTDLTGTIQKISRDTHGRV 1015
D 986 QGEILDRREYVSTGSLKSEKTSAGKTYDMFYSSYQGLLSTYDVLQGEQVNRXYDTCSRL 1045
QY 1016 TQIYDSSIKTILAND---DLNRHIGSQVTLATLGHMLTTTVEPRGLNREIRGLKCSG 1071
D 1046 BETRLGEVYSTFADVDFGR.LAKITTTQGTDSRRRTYALISLEYDRLREVGRFP-DIDG 1104
QY 1072 HTLDIQSMLTKQOLANRIVK.LNGVLQRTQOYSYDRNRLLQYKCDGACETLDKXHSIV 1131
D 1105 VEQGVQVYDDVDVQRTLSGTEILADENHNVDTSRLTQYQCTGQRVPDYPGRLL 1164
QY 1132 TQNTFYDIYGNITACHTTFADGTEHATFKFANPTDPCQLTEVHNT-HPDMPDNIRLKYD 1190
D 1165 GQNFIFDQANMLTFYTRF-DQGSNNARY-FYEGDDPVQLSVTNTNAYYPPPEINILAYD 1222
QY 1191 KAGRININTDNHNGENTFYDTLGRLOV-----GQSSVGYDPLANLVSOQKDTLDCEL-X 1245
D 1223 PDGNTL-DIDAGRT-LKYDPLGR.LIEVGTLSAGIHQYDPODLTBE-TRGGGRDLRF 1277
QY 1246 YRETLVNEVRNGEMIRLLRTGETI.IAQORASKVLLTGTDSQGSVILTSDKONLSQEAYS 1305
```

```
D 1278 YRDGEILANQSGSGGQSTFRKRDYLLAEQGDRTLLATASDSVLDGVAADVNRWYT 1337
QY 1306 AYGHKSTANDASIIIGYNERADPVSGVTHLNGVSYSDPTLWRFHTPDLSFPFGGIN 1365
D 1338 AYGHASGDEPPHGRILDFNGELSEADTGOMQLNGYRAYSVLWRFPNSPDSWSPBGGMN 1397
QY 1366 PYSYCLGDPINRSDPSGHL.SWQAMTGI.GMGIAGLL-----TATGMAI.AAAGIA 1417
D 1398 AYAYVGEDPVNMEDPFGH-----FGLFTPPKLYRLAKTTPTLTTSRGVEKIPGL 1449
QY 1418 AAIASSTTALAFALSVTSIDTISI-----VSCALBPASPRASITL 1459
D 1450 TSGRKEVTKLSKITMDMDDLTDVAESYPARISAKIDNKPGIAATLTKQTEAKDALN 1509
QY 1460 WVMGMCAGL---AESATK 1476
D 1510 YLSAHVNPQGITAHARRAK 1529
```

## RESULT 4

```
Q883V9_PSESM
ID Q883V9_PSESM PRELIMINARY; PRT; 1562 AA.
AC Q883V9;
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE YD repeat protein.
GN OrderedLocustNames=PSPTO2239;
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=323;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=DC3000;
RX MEDLINE=2834015; PubMed=12928499; DOI=10.1073/pnas.1731982100;
RA Bueli C.R., Joardar V., Lindeberg M., Selenyut J., Paulsen I.T.,
RA Gwinn M.L., Dodson R.J., DeBoy R.T., Durkin A.S., Kolony J.F.,
RA Madupu R., Daugherty S.C., Brinkac L.M., Beaman M.J., Haft D.H.,
RA Nelson W.C., Davidson T.M., Zafar N., Zhou L., Liu J., Yuan Q.,
RA Khouli H.M., Fedorova N.B., Tran B., Russell D., Berry K.J.,
RA Utechtack T.R., Van Aken S.E., Feldblyum T.V., D'Ascenzo M.,
RA Deng W.-L., Ramos A.R., Alfano J.R., Cartinhour S., Chatterjee A.K.,
RA Delaney T.P., Lazarowitz S.G., Martin G.B., Schneider D.J., Tang X.,
RA Bender C.L., White O., Fraser C.M., Collmer A.;
RT "The complete genome sequence of the Arabidopsis and tomato pathogen
Pseudomonas syringae pv. tomato DC3000."
RL Proc. Natl. Acad. Sci. U.S.A. 100:10181-10186(2003).
DR EMBL; AB016853; AAO55755.1; -; Genomic_DNA.
DR TIGR; PSPTO2239; -.
DR InterPro; IPR006530; YD.
DR Pfam; PF05593; Rhs_repeat_6.
DR TIGRFAMs; TIGR01643; YD_repeat_2x; 3.
KW Complete proteome.
SQ
SEQUENCE 1562 AA; 175713 MW; 8DC10DA1BEF37BF1 CRC64;
```

```
Query March 18.2%; Score 1614; DB 2; Length 1562;
Best Local Similarity 30.3%; Pred. No. 1.2e-76;
Matches 490; Conservative 239; Mismatches 700; Indels 190; Gaps 51;

QY 14 MSDNFEFTQANNTSAVSGGVDPRTGLYNIQITLGHIVGNGNLGPTLPLTLYSPPLNKT 73
D 1 MTSSTVSHSNANFNMSYLGSGVDPRTOGYTIVSLPEVKSNGLGRFPVPLVLSNPLAVQ 60
QY 74 DIGFGIFNGLSVYDRKNSILSTGBNRYVT-ETDKTVK.LQOKLDNLNRFKDKKNC 132
D 61 DSGFGGLMNLQLSQYDGTIRIVLSIGSGTFFKESLSLGDQLMMPKXKDSFFHYXQ-DTR 119
QY 133 YRIHKGSDIEVLGFPNNMFDLKVPRKLLNPAGAHATYIDNFEATQPLRIRIYDLDLGH 192
D 120 YRVVHKSQGVLEVLDSLGNRIALPVRIYSPBGHGLTHYASFGAYOMLSBVDD-DGQ 178
```

QY 193 DIPLANLEYOGLIKITLITLFPQGEKGYRTBLRPLNRQNLSTHNSFLGNEPLTWSFGTTP 252  
 DB 179 VILITTD-----SISVALLLYGAPKADAEFWMILSGSNRVARIELPTAKASRFTYTSI 234  
 QY 253 IKGKGLIGWITMTABGKLETNYVSNNOGHHFPOSANL.PVL.PYTLKQVAGACOPA 312  
 DB 235 IKGHSC-----IASVDTFVGHEVDYFQDS--GHQFPLISAGREPLPRVTRHLTTGFLQPE 288  
 QY 313 IOAEYSI-----TSHNVVGGSGNGIW--NNKLDNLYGLM--TEVYGGTESRRYDQKGDH 364  
 DB 289 VDRVYAKKQAGREBNPLGAGLIDJAMEDNGLNLYRIGAPYLSSTETELRVNDV--- 345  
 QY 365 QIVARETANNVHLLTSECKOONGYIQTETAYAIIGHNDPSQFOLPKTETMR- 423  
 DB 346 -VASIEHVPQFHLAETTRQNLSTLEVDRTYIEEGKPPDLPQNCOLPKEXRTTMR 404  
 QY 424 SADNSY-RSEITETTFPESGNPLTKVIKDKTKQIKIISPTHEMYYPAGEVNDCPPEY 482  
 DB 405 SPDSVPRTEIVSDYDYGNLAQTOANGVTES-----EYVSQGE-DGCPDPDG 456  
 QY 483 PTRPVKIIQTPYSEKDEKFIQYRY---SLISQSHVTLKIF---ERHYSATQ 533  
 DB 457 FVRLTKAKSVVPAQSDYGHALVLTARYKALPALASGQNLMLAASETLLQTTDGEK 516  
 QY 534 LNSTLFOYNTDKS-----BLGRLLKOTECTGKENGKT---YSVYHKEFTYTKQDDTLQSH 586  
 DB 517 ELQOTTYYIEDNYDAFOYGRIRHQSVTLEGLSTTDYRYDLQDDP---DQTVQTV 573  
 QY 587 SITTHDNFTIHRQVR---SRYTGRLEPSDITDKDITVOMSYDKLGRLLTTLN--SGTPYA 642  
 DB 574 QIVT--GFDMQKVIYRLEHSLFTGEPILNRDNDVEIRYDYNLRVYSEIVSNKEBK 631  
 QY 643 NTLTYBELANLQDNRPPVITTTDVNGNQLRNEFGAGHNSQCLKSDGDKGKPTTIH 702  
 DB 632 ATRHYEQOLAKVTKDOAEORLF---DYKNVQYTSRFPGLGVIYEABADNDPVRHRLD 688  
 QY 703 TOQ-----YDEQGHHTSTYSDYLTNGRQOTDPKVMLSMSKSYDNMGQANTHMSYGS 757  
 DB 689 LRQYEAAYDAMGDKVETSYDML-----DOOKALITYPEYDDDOOLSTVGPGCVT 741  
 QY 758 EKITVDITLTA-----KOLQSNANNVQGEKVTY--TPSQOPIQITLPEACHLOSCH 811  
 DB 742 TIEQDVEYTOASNGPIORRWTESNDGLQTSSEVETMLNLFDEPTRSVRLDRDMLSBPV 801  
 QY 812 TLTR---DGMDRVAKETDAI---GQCTTYQYDNNRYIQTLPRTTYVNRKAPFSTDTL 865  
 DB 802 SLRHYOYDGLRVLKEVSGLPPIRSTYTYGDVDRVANTLPDGAVRRRYAPHSGBDL 861  
 QY 866 ITDIRV--NGIS--LGOOTPDGLSRLOSQDQGRVMAVYTSAGNDQCPSTVTPDQGFH 921  
 DB 862 PAWIGVHNGKASVLSGOKTDLRITVSTTSGERELSTSDLMQ--PKTVKLPBSGQID 920  
 QY 922 YOYOPELDVAVLQ-----VASNEITQOFSYNPATGALLKAVAGESGLTPIYPSGRKLM 975  
 DB 921 YDVLPELGDEPLKKTQSDVYARLKLTDYTTDPQANALISSBGELOREHYSTGLKS 980  
 QY 976 ENINDM--KKMSYLMTLRLGLENCTYDULGTIQKISDTHGRVTOIKOSSIKTTANTYDL 1032  
 DB 981 EORTSGOIEYTMHRYSRGLPLSYIIVLQGOEQLSVYDDGRLAQTSIGREVSPFTYDTF 1040  
 QY 1033 NRHIGSOVTLATGMHMLTTTVEPDLNREIGRKLCDSSGHTLDDIOGMLKQOLANRYK 1092  
 DB 1041 GRPASITLDDSSNGOVVISLEVYAGREAKRT--TINGANQOMVQYDDVDQVVKETLS 1099  
 QY 1093 LNVGLQTEQYSDSRNRLNQYKCDACEPTKXGHSIVTQNTYDYIGNITACHTTPAD 1152  
 DB 1100 EGVAYIIEBHXYGLDQRLTQYDCSGKORPVDYGMTISQVFFDGLNMLTLYTTTF-D 1158  
 QY 1153 GTBEHAFKXANPDPQOLTEVYHHTH-----PMPNRIKLKTKAKAGVIN 1197  
 DB 1159 GGRRRAY--FYEIGDPAQLTRVNTQOLAMVNAARLIPVKNKDSITYPBIIRLTYPDGML-- 1215  
 QY 1198 ITDNGHNTENTYDTLRL---QNGQGSV--XGYDPLNRLVSOQKTDLDCELYRETMVLN 1253

DB 1216 ITDEADL--LSYDPLGRLLBVMMPAGDVYRRDPQORLAGTGE-----QRFPRDVLAS 1269  
 QY 1254 EVRNGEMIRLLRTGETIITIAOORASKVLLTGTDSQOQVILTSDKONLSQEAYSAVKGKST 1313  
 DB 1270 QLGASQSTYMRGQGYLLAEQCSDDLFTSISNSVLSVHHPDGVNRSRTYVGHSSGD 1329  
 QY 1314 ANDASILGNVGERADPVSGYTHLGNGRSYDPTLMRTHITDLSLSPFAGGINPYCYCLD 1373  
 DB 1330 DPPAGRGYNGELHEHTDQWOLLGNGYRAANPVMRHSBDSWSPFEEGGIINAAVYEGD 1389  
 QY 1374 PINRDPESG-----LSWQATG-----IGMGIALLLTITATG 1407  
 DB 1390 SVNGVDPPTGCMFRLPRTFRTRLEKQAMRADKKEFRLIEODIKNEGRLGRYAQARLD 1449  
 QY 1408 MAIAAGG---IAAIASTSTTALAFAALSVTSDITSVSGALEDA--PKASILGWVS 1462  
 DB 1450 OAKSARKNKAYISLSLVGRREKALE--AFTLKANTGVWSSRSBSITPKAEASGSGIS 1507  
 QY 1463 MGKAGLAASAIKGTAKLATHGAPADEBNALIKSTSSSRKMKVTSLSPREIYRN 1521  
 DB 1508 -----NLGFGSGSDRDVGLRVSSA---KSEVLARQD--FVRN 1540  
 RESULT 5  
 Q7NMU2\_CHRVO  
 ID Q7NMU2\_CHRVO PRELIMINARY, PRT, 1385 AA.  
 AC Q7NMU2;  
 DT 01-MAR-2004 (TREMREL, 26, Created)  
 DT 01-MAR-2004 (TREMREL, 26, Last sequence update)  
 DT 01-MAR-2004 (TREMREL, 26, Last annotation update)  
 DE Nematocidal protein.  
 GN Name=Thsa; OrderedLocusNames=CV11897;  
 OS Chromobacterium violaceum.  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
 OC Neisseriaceae; Chromobacterium.  
 NCBI\_TaxID=536;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=ATCC 12472 / DSM 30191;  
 RX MEDLINE=22882880; PubMed=14500782; DOI=10.1073/pnas.1832124100;  
 RA Vasconcelos A.T.R., de Almeida D.F., Hungria M., Guimaraes C.T.,  
 RA Antonio R.S.V., Almeida F.C., de Almeida L.G.P., de Almeida R.,  
 RA Alves-Gomes J.A., Andrade E.M., Aratijo U., de Araujo M.F.F.,  
 RA Abouli-Filho S., Azevedo V., Baptista A.U., Batus L.A.M.,  
 RA Batista J.S., Belo A., van den Berg C., Bogo M., Bonatto S.,  
 RA Bordignon J., Brito C.A., Brocchi M., Burtly H.A.,  
 RA Camargo A.A., Cardoso D.D.P., Carneiro N.P., Carraro D.M.,  
 RA Carvalho C.M.B., Cascado J.C.M., Cavada B.S., Cuelire L.M.O.,  
 RA Creczynski-Pasa T.B., Cunha-Junior N.C., Fagundes N., Falcao C.L.,  
 RA Fantiuati F., Farias I.P., Felipe N.S.S., Ferrati L.P., Ferro J.A.,  
 RA Ferro M.I.T., Franco G.R., Freitas N.S.A., Furlan L.R.,  
 RA Gazzinelli R.T., Gomes B.A., Goncalves P.R., Grangeiro T.B.,  
 RA Gracispaglia D., Gilsard E.C., Hanna E.S., Jardim S.N., Laurino J.,  
 RA Iacoi L.C.T., Lima L.F.A., Loureiro M.F., Lyra M.C.C.P.,  
 RA Madeira H.M.F., Manfio G.P., Maranhao A.O., Martins W.S.,  
 RA di Mauro S.M.Z., de Medeiros S.R.B., Meisner R.V., Moreira M.A.M.,  
 RA Nascimento F.F., Nicolas M.F., Oliveira J.G., Oliveira S.C.,  
 RA Paixao R.F.C., Parente J.A., Pedrosa F.O., Pena S.D.J., Pereira J.O.,  
 RA Pereira M., Pinto L.S.R.C., Pinto L.S., Porto J.I.R., Potrich D.P.,  
 RA Ramalho-Neto C.B., Reis A.M.M., Rigo L.U., Rondinelli E.,  
 RA Santos E.B.P., Santos F.R., Schneider M.P.C., Senarez H.N.,  
 RA Silva A.M.R., da Silva A.L.C., Silva D.W., Silva R., Simoes I.C.,  
 RA Simon D., Soares C.M.A., Soares R.B.A., Souza E.M., Souza R.L.L.,  
 RA Souza R.C., Steffens M.B.R., Steindel M., Teixeira S.R., Umenyi T.,  
 RA Vettore A., Wasseem R., Zaha A., Simpson A.J.G.;  
 RT "The complete genome sequence of Chromobacterium violaceum reveals  
 RT remarkable and exploitable bacterial adaptability".  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:11660-11665(2003).  
 DR EMBL; AB016916; AAQ59561.2; -; genomic\_DNA.  
 DR InterPro; IPR006530; YD.  
 DR Pfam; PF05593; Rhs repeat; 6.  
 DR TIGRfam; TIGR01643; YD\_repeat\_2x; 3.





Best Local Similarity 44.6%; Pred. No. 8,66-76;  
Matches 342; Conservative 119; Mismatches 275; Indels 30; Gaps 17;

```

QY 741 YDNMGQJANTHMSGVSEKIVTDVITLTATKQLOSNNSNVQTEKVTYTPSQOPIQITL 800
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 4 YDSMGQNHLLTVFSDGERSVYDPIITRRAT--LQPEGSQKLGQQLTEVNIAGLPKIVTQ 61
QY 801 FDEAGHLOSCHTLTRDGMVRKRETDALGQCTIYQYDYNRVVIOITLPDGTIYNRKYAPF 860
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 62 YDSGCTQGSANHYVDGLGQARKETDELQVTLLEYDHFGRVVOTLPENTIIQKSYAPH 121
QY 861 STDITLDIRVNGISLQGOQTFDGLSRLTQSDGGRVWATYTSAGNDQCESTVITPDGQFI 920
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 122 STASLTIGISVNNFSMGNTQFDSLERLTETTSGRTSAFSYENAS-SVPAVATPAGTAV 180
QY 921 HYQYQPELDVAIVQVANSNEITQOFSYNPVYALAKAV-AEGOSLPIYYPGRLKME-NI 978
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 SYEYLKEIAGNAVKKISAPEILKTWDYDALTGAMTSATQAGMIRQMTYTPSGLLKNETSM 240
QY 979 ND--MKMSYLTMLRGLNGVYDITLGTIOKISRDPHGRVTOIKDSSIKTTLANYDDLNRH 1035
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 PDGAAGKSTATYTLAAGPOSTYDVGTVQRYDYDDEHRRIGIEDNDIKVSLGIDAFGRF 300
QY 1036 IGSQVTLATGMLTTVEFPDGLNREIGRKLCDSSGHTLDIQSMWLTQOLANRIYKLG 1095
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 TKQQATDKKTGAVLSTLTLYDDLNRREIKREISAGQSGLVIEQTYQRHLLKERITQGR 360
QY 1096 VLOQTEQSYDSRRNLNOYKCDGACPTDKYGHISVTONFTYDIYGNITACHTTPADGTE 1155
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 361 TTLKEVFAYDSRRNRLIEYTCNGEARPDQPYGKAIHQFTSYDALGNMTKTQTFDSGQ-R 419
QY 1156 DHATFKFANPTDPCQLTVEHHHTHPDMPDNIRLKYDKAGRVINITDNGHNTENFTYDLGR 1215
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 420 NTAATYIY-SADPQOLKVNNDHSDYKPEITLEYDKAGMR--DEAGRT--LRYDALGR 474
QY 1216 LQ--NG--QGSVYGYDPLNRLVSQKT-DTLDCELYYRETMVNEVR--NGEMIRLLRTG 1267
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 475 LQGVNAGAKGQGYAYDALNTLVSVQVDEPIYDLYRADDLVGEARRDSSQTRYVXSN 534
QY 1268 ETTIAQ--QRASKV-LLTGDSQOSVILTSDKONLSQEA-YSAYGKHKSTANDASTILGY 1322
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 535 GCCCGQCTKGSSNTSRLTTNNQOSVLSVSEGNHAPQDCITYRGRTPQTEPSTVLGF 594
QY 1323 NGERADPVSGVTHLNGYRSYDPTLMRFHTPDSLSPFGAGGINPYSCLDGPINRSDPSG 1382
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 595 NGERLDPVSGTYHLNGYRAYNPILMRFCPSDMSWSPGAGINPYAICDGPINRVDPN 654
QY 1383 HLSWQAMTGMGIAGLLTATAGMAIAAAGIAAIAASTSTTALAAGALSTSDITSI 1442
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 655 HLSWQABELGIGLVGLVLAFTAGTSIAAAGIASAIESASISLVVGTGLVADVASI 714
QY 1443 VSGALBDSAPKASSILGVWSMGGAAGLAE--SAIKGTGKLAATHL 1485
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 715 ASGALBDBANQASATLGWISLGLGPGAVSGLATPAAAGKLLISGL 760

```

## RESULT 8

052880 COXBU PRELIMINARY; PRT; 774 AA.  
AC 052880;  
DT 01-JUN-1998 (Tremblrel. 06, Created)  
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
DE Hypothetical protein orf 774.  
GN Name-orf 774;  
OS Coxiella burnetii.  
OC Plasmid QRS.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;  
OC Coxiellaceae; Coxiella.  
NCBI\_TaxID=777;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RA Lautenschlaeger S., Jaeger C., Willems H., Baljer G.;

RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Y15898; CA475841.1; -, Genomic DNA.  
DR InterPro; IPR006530; YD.  
DR Pfam; PF05593; RMS repeat; 6.  
DR TIGRFam; TIGR01645; YD\_repeat\_2x; 2.  
KW Hypothetical protein; Plasmid.  
SQ SEQUENCE 774 AA; 83774 MW; 918A543D7991BD8 CRC64;

Query Match 17.2%; Score 1530.5; DB 2; Length 774;  
Best Local Similarity 44.5%; Pred. No. 1,46-74;  
Matches 341; Conservative 120; Mismatches 275; Indels 31; Gaps 18;

```

QY 741 YDNMGQJANTHMSGVSEKIVTDVITLTATKQLOSNNSNVQTEKVTYTPSQOPIQITL 800
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 4 YDSMGQNHLLTVFSDGERSVYDPIITRRAT--LQPEGSQKLGQQLTEVNIAGLPKIVTQ 61
QY 801 FDEA-GHLOSCHTLTRDGMVRKRETDALGQCTIYQYDYNRVVIOITLPDGTIYNRKYAP 859
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 62 YDSNGTEQGSANHYVDGLGQARKETDELQVTLLEYDHFGRVVOTLPENTIIQKSYAP 121
QY 860 FSTDITLDIRVNGISLQGOQTFDGLSRLTQSDGGRVWATYTSAGNDQCESTVITPDGQFI 919
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 122 HSTASLTIGISVNNFSMGNTQFDSLERLTETTSGRTSAFSYENAS-SVPAVATPAGTAV 180
QY 920 IHYQYQPELDVAIVQVANSNEITQOFSYNPVYALAKAV-AEGOSLPIYYPGRLKME-N 977
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 VSEYELKEIAGNAVKKISAPEILKTWDYDALTGAMTSATQAGMIRQMTYTPSGLLKNETS 240
QY 978 IND--MKMSYLTMLRGLNGVYDITLGTIOKISRDPHGRVTOIKDSSIKTTLANYDDLNR 1034
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 MPDGAAGKSTATYTLAAGPOSTYDVGTVQRYDYDDEHRRIGIEDNDIKVSLGIDAFGR 300
QY 1035 HIGSQVTLATGMLTTVEFPDGLNREIGRKLCDSSGHTLDIQSMWLTQOLANRIYKLN 1094
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 FTQQATDKKTGAVLSTLTLYDDLNRREIKREISAGQSGLVIEQTYQRHLLKERITQGR 360
QY 1095 GVLQTEQSYDSRRNLNOYKCDGACPTDKYGHISVTONFTYDIYGNITACHTTPADGT 1154
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 361 RTTLKEVFAYDSRRNRLIEYTCNGEARPDQPYGKAIHQFTSYDALGNMTKTQTFDSGQ- 419
QY 1155 EDHATFKFANPTDPCQLTVEHHHTHPDMPDNIRLKYDKAGRVINITDNGHNTENFTYDLG 1214
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 420 RNTATYIY-SADPQOLKVNNDHSDYKPEITLEYDKAGMR--DEAGRT--LRYDALG 474
QY 1215 RLQ--NG--QGSVYGYDPLNRLVSQKT-DTLDCELYYRETMVNEVR--NGEMIRLLRT 1266
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 475 RLQGVNAGAKGQGYAYDALNTLVSVQVDEPIYDLYRADDLVGEARRDSSQTRYVX 534
QY 1267 GETTIAQ--QRASKV-LLTGDSQOSVILTSDKONLSQEA-YSAYGKHKSTANDASTILG 1321
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 535 NGCCVGQCTKGSSNTSRLTTNNQOSVLSVSEGNHAPQDCITYRGRTPQTEPSTVLGF 594
QY 1322 YNGERADPVSGVTHLNGYRSYDPTLMRFHTPDSLSPFGAGGINPYSCLDGPINRSDPS 1381
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 595 NGERLDPVSGTYHLNGYRAYNPILMRFCPSDMSWSPGAGINPYAICDGPINRVDPN 654
QY 1382 GHLSWQAMTGMGIAGLLTATAGMAIAAAGIAAIAASTSTTALAAGALSTSDITS 1441
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 655 GHLSQABELGIGLVGLVLAFTAGTSIAAAGIASAIESASISLVVGTGLVADVAS 714
QY 1442 IVSGALBDSAPKASSILGVWSMGGAAGLAE--SAIKGTGKLAATHL 1485
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 715 IASGALBDBANQASATLGWISLGLGPGAVSGLATPAAAGKLLISGL 761

```

## RESULT 9

042079\_PSESY PRELIMINARY; PRT; 1682 AA.  
AC 042079;  
DT 13-SEP-2005 (Tremblrel. 31, Created)  
DT 13-SEP-2005 (Tremblrel. 31, Last sequence update)  
DT 13-SEP-2005 (Tremblrel. 31, Last annotation update)  
DE YD repeat.

GN ORFNames=Payr\_2040;  
 OS Pseudomonas syringae pv. syringae 8728a.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; Pseudomonas.  
 OX NCBI\_TaxID=205918;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=8728a;  
 RG DOE Joint Genome Institute;  
 RA Chain P., Latimer F., DiBartolo G., Copeland A., Lykidis A., Trong S.,  
 RA Nolan M., Golden E., Thiel J., Malfatti S., Lapidus A., Deter J.C.,  
 RA Land M., Richardson P.M., Kyriides N.C., Ivanova N.;  
 RT "Comparison of two complete genome sequences of Pseudomonas syringae  
 RT pv. syringae 8728a and pv. tomato DC3000.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(2005).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=8728a;  
 RA Loper J.;  
 RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=8728a;  
 RA Fell H., Fell W.S., Lindow S.E.;  
 RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: CP000075; AAY37083.1; -; Genomic DNA.  
 SO SEQUENCE 1682 AA; 186410 MW; CFPF95C541B81ECF CRC64;  
  
 Query Match 16.8%; Score 1490.5; DB 2; Length 1682;  
 Best Local Similarity 29.0%; Pred. No. 7.4e-72;  
 Matches 477; Conservative 237; Mismatches 715; Indels 217; Gaps 53;  
  
 QY 14 MSDNNEFTQANNNTSAVSGVDPRTGLVNIQTILGHIVGNNGPLPLTLYSPANKT 73  
 DB 1 MAASTVSHANLNMWCLCKSGVDPRTGLVNISSLPLQSDNLRGPGRDLSSQLNTL 60  
 QY 74 DIGGIGFNGFGLSVYDRKNSLLSTGENYKVIETDKT--VKLOQKLDMLRFEKDKEN 131  
 DB 61 DSGVGLGMNQLDSYDPATQILSLSTGETRVVDTSSNGQLTMREKLDTFHFVK-IDDE 119  
 QY 132 CYRIHKSGLIEVLTFGNNAFPLKVPKLLNPAGHAIYIDWN-FEATQPLNRIYDDL 190  
 DB 120 SYRAHSGGLVEILHLSGKMKAMPVKIIAPSGHSIALKHTVFNSTYRLASITDDL 179  
 QY 191 GHQIPLNLEYGGLIKTI-----LTLFPGQKGYTELRFLNKLQNS--IHNFLGN 240  
 DB 180 -----QTLTKIARSTSVYLDLHPDEGTGTPPLARFMTTLAGSOKRVSRITLPT 228  
 QY 241 ENPLTMSFGYTPIGKNGILGOWITSMTPAPGLKETYVYNNNOGHFPOSANLPLVPLVT 300  
 DB 229 ENKASWRFEGLENGNOLC---VAHVETPAGSSENVY--QDEGHAFPSHAGRPVPRVT 283  
 QY 301 LMKQVPAGAPAIQAEYSY-----TSHNVYGGSGNGIW-NNKLDNLVGLMTERVYGSTES 354  
 DB 284 KHVYDPLGNLAKIDVRYTYKGGQORSRNPLGAGLPIMEDNGLNLKYLYODVYVCTES 343  
 QY 355 RRYKDXKHQOIVRIETNNYHLJTECKQONQYIOTTEYAYAIIGHNFDSPSPOL 414  
 DB 344 LWDVNRK---AVSIEKTFNRFHLQTLLEVTTQNNQLVTTAYVNILOGEHYVSRQPNCOL 399  
 QY 415 PKTKETETRSAD--NSYRSEITETTFPDSGNPLTKVKDKKTKQKIISPTHWEXYPAGE 472  
 DB 400 PSEITTRMQLDAERTRTEVTDVSYGNLVHTPADG-----IEVSSW--YPAAGG 452  
 QY 473 VDNCPPEPVFTFVKKIQTQPYDSEKDDPEKTIQYRSLIGS-----OSHV 520  
 DB 453 -DCCPDADAEFVSLKEKVKVPASSQLGAPTLATRYRQTLPALADSELPDWIVPESRT 511  
 QY 521 TLKIEBHRYATQULNSTLFOY--NTDKSEL--GRLLTQTECTYGENKNTSYVHKFTYTK 577  
 DB 512 LLDQESD--GTCELQVLELYINQPDAPFLHGRGTGNISLNNNDVTV---YVESKSK 566  
 QY 578 --ODDTLQOHSITV--HDNFTIHSQVRSRYTGRLEPDTDKIDVITQMSYDKLGRLLTR 633

DB 567 SRQLEVVQULIDITTTGFDNAKISKQOQSILTQVLL-TLESGVEIRRYVDVLRITKE 625  
 QY 634 TLNSGTPYANTLTYVEYELNLIQDNRPEVYITTTVDNGNOLRNEFDGAG-----HVS 686  
 DB 626 TIAPDSSDEASRELYLTLCSAAGQAEQVIV---DARKVTRSVLDJGRLIIEERDHI- 681  
 QY 687 QCLKSDGDKKFTTIHQVDEGRHHTSVSYDLTNGRGOTPDKXHLMSKSYDNMGQ 746  
 DB 682 ----DSNDLOAMRIIHTAQYANNNVOHERDYMPAN--QO---RATNTAYIDMNV 731  
 QY 747 IANTHWSYGVSEKITVDP-----TLTATKOLQSNVNVQKEXETVYT---PSCQP 795  
 DB 732 QCVTISHGQVTHQVYDPIGNEHKGVPQKWTQSSDAEPLIGRSSTWLMNGKPKIK 791  
 QY 796 IQITLPEAGHL-QSCHTLTRDGMDRVREKTDAGCCTIYQYNNRVIOITLPDGTIVN 854  
 DB 792 IKIKTONADGKQSGQTFLYDGLGRCTEQTDSNHTLPSYDAMSRYVTRRLPDGSEAH 851  
 QY 855 RKYAPSTDLIDIRV---NGIS---LGGQTFDGLSRITQSDGGRVMAVYTSAGNDOC 908  
 DB 852 RQYAPHSNTELPGLKMMHSDGLTNTLAGKQAFDGLRLTLSQAGRTIEQYEDGRQV 911  
 QY 909 PSTVITPDGQFIHYOYQPELIDAVLQVANSBITQOFSYNPVTGALLKAVAEQSLTPIY 968  
 DB 912 -KTRKTAKGDISYTNALATDQIVSSTAPDVTGFGVDVNSARLTSATNQGGRVAYD 970  
 QY 969 PSGRLKKNENIMDK---KMSYMTLRGLNGYTD-----LTGITQKI 1007  
 DB 971 AHQUTKETWKDLOGRTWKTVHRTSLQGRMYRTDQKDVNSVKGKVGKRIKGVKGV 1030  
 QY 1008 SR-----DTHGRVYQIDSSIKITLNVDDLNRIHGSQVTDLATGMLTTPYEPGLNKEIG 1063  
 DB 1031 ETTIRYDDFERLENVQAGNVEVITLVDLQPSBMTTCDDAAGRLNRRKKTIDQGGEL 1090  
 QY 1064 RKLDSGHTLIDIQSGWLTQOLANRIVKLVGLVQRTQYSYDSRNRLNKKYCDGACPT 1123  
 DB 1091 RTQIAGNHPRITLQOQMDQDGLMQSRHLQEAQNTLLETFKYDARGRLTVIYSGSLPV 1150  
 QY 1124 DKYGHSLVTONFTYDIGNITACHTTFADGTEDHATFKFANPTDP-----COLTVHH 1176  
 DB 1151 HASGRALIKQTFPDSLDNMIQTTTEPADASTEMAFPHYKEDDPDLSDRCOLLGITY 1210  
 QY 1177 THPDMQNIKLYKDKXGRVININDNGENTENFYDTLGRL---QNGG---SVYGYPLN 1230  
 DB 1211 RPRGTPDPFTRFYDANGN--QLDDEHGN--RLYYDSQRLRYEKPAGBPISTAYADGHD 1266  
 QY 1231 RLVSOKTDLIDCEL--YYRETMVNEVRNGEMTRILRTGETIIAQO---RASKVLLTGT 1285  
 DB 1267 HLAFTTNGS--DSRLARFYQDQQLSSTVQDORRTQFLYIDQPVGQQTIGSPATILLTD 1325  
 QY 1286 SQQSUILTSQKUNLSQEAAYSAYGKHSTANDASILGNGERADPVSGVTHLNGYRSYDP 1345  
 DB 1326 ANOSVLAFFQDDKLRARVAYAYGERHSDDALLVAGANGEICEKDTGMWYLLGNGYRAYNP 1385  
 QY 1346 TLMREHTPDSLSPGAGINPVSYCLGDPINRSDPSH----- 1383  
 DB 1386 GMRKFPSPDLSLSPGAGVAPYTYCLGNPLAMRDPGTHDASSQSGRLRRPDEDAIPAE 1445  
 QY 1384 --LSWQAMTIGIGI---AGLLTLTATGMAIAAGGIAAALASTSTTA----- 1427  
 DB 1446 GGMGIQIOWIMLANGIWTAIVGAATVAFSG--LAAQSVAVTVLGMNTAQTLAAVATGL 1503  
 QY 1428 LAFGALSVTSDITSIVSGALE--DASPKASIIIGWWSMGKG-AAAGLAESAIIKGTKLATH 1484  
 DB 1504 LAAGTVLGAASTAATYTAIVGAVKGETAFRLQELVLAVALPIDIAGVFRVSAIKAAAS- 1562  
 QY 1485 LGAFADGENALIKSTSE--SSRIKMG 1509  
 DB 1563 -----KASSRVSSETASRVSVG 1579

RESULT 10



Q88LP7\_PSEBK  
ID Q88LP7\_PSEBK PRELIMINARY; PRT; 1632 AA.  
AC Q88LP7;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
GN OrderedLocustNames=P1882;  
OS Pseudomonas putida (strain KT2440).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=160488;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22423060; PubMed=12534463;  
DOI=10.1046/j.1462-2920.2002.00366.x;  
RA Nelson K.B., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,  
Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,  
Brinkac L.M., Beanan M.J., DeBoy R.T., Daugherty S.C., Kolonay J.F.,  
Madupu R., Nelson W.C., White O., Peterson J.D., Kouri H.M.,  
Hance I., Chris Lee P., Holtzapple E.K., Scanlan D., Tran K.,  
Morzzer A., Utterback T.R., Rizzo M., Lee K., Kosack D., Mestl D.,  
Weiler H., Luder J., Scjepandic D., Honeisel J., Straetz M., Heim S.,  
Kewitz C., Bisen J.A., Timmis K.N., Duesthoeft A., Thiemler B.,  
Fraser C.M.;  
RA "Complete genome sequence and comparative analysis of the  
RT metabolically versatile Pseudomonas putida KT2440.";  
RL Environ. Microbiol. 4:799-808(2002).  
DR EMBL; AE016780; AAN67501.1; -; Genomic\_DNA.  
DR TIGR; PF1882; -;  
DR InterPro; IPR006530; YD.  
DR Pfam; PF05593; RBS repeat; 7.  
DR TIGRfam; TIGR01643; YD repeat 2x; 4.  
KW Complete proteome: Hypothetical protein.  
SQ SEQUENCE 1632 AA; 182290 MW; F07C2A5820160401 CRC64;  
Query Match 15.9%; Score 1416; DB 2; Length 1632;  
Best Local Similarity 28.4%; Pred. No. 8e-68;  
Matches 472; Conservative 232; Mismatches 734; Indels 226; Gaps 52;  
QY 14 MSDNNEFTQANNPFSAVSGVDPTGILNITLGHVGNNGNIGPTPLPLTSLSPINKT 73  
DB 1 MSTSSLSHNSLNFSPWESGVDSTGQYTLSTIKLPGLGANYFSGPELISLNFPLNMI 60  
QY 74 DIGIGFNFGISYDRKNSLSLSTGENYKVI-ETDKTVLKQOQKLDNLFEKDLK--- 129  
DB 61 DSGMGKGNLSLTQPAHPTQVITTYSGSFVYTGSSGRLQMOEKLHFFHYEPAPRG 120  
QY 130 -ENCYRIHKSGLIEVLTFGNNAFDLKVPKKLNPAHAITYIDW-PEATQPLNRIYD 187  
DB 121 GNARYVVRHSGVLVEIEMSGANGRIMALPEIYATGHRIDLQYQPNSSYMLLSISD 180  
QY 188 -----DIDGHIPLNLEYOGLKITILTFPGQEGYRTFLRNQLNS--INNF 236  
DB 181 EREBILIEIENDSDSIELREBPYQ-----DNGQPVALYANN-LITDWTSTI 227  
QY 237 SLGNENPLTWSFGYPIGNGILGQWITSMTAPGAKETVYNNNOGHFPOSANLPLV 296  
DB 228 VLPHIELASWRLVLY--NGLL--CVSKVERPTGAREVLY--QDRHLTFPGDAR-PPL 280  
QY 297 FYVTLMKQVPGAGQPAIDAEYSYTS-HNYVGGSGNGIV-NNKLDNLY-GLMTEYNGSTE 353  
DB 281 PRVTHVIEPRGGQAFQRTTYTPGLNPNFLYGAGIGMSDNGLDMLYESKRYDEYQYVE 340  
QY 354 SRRYDKGHDQIVALEERTYNNYHLLTSECKQONGYIQTFAYVAIIGHNDSQSPQFO 413  
DB 341 TLR--DEGG-TALRDIPTFNRFHLITSTRYVONNCHEV-TWQYNIDVPPNQQVSTLQ 396  
QY 414 LPTKTEITWRSADNGYSR--EITETTFDESGLPLTKVIVKDKTKOIKISPTHEWEYYPAG 471  
DB 357 MPRIQGTKRLADNITRSLLETVEITTYDSSGNIIRKLANGYTER-----QEWYGTQA 449  
QY 472 EVDNCPPEPYGFTFRVKKIIQTTPYDSEFKDPEKFIQYRYSLLIG--SQSHVTLK----- 523

DB 450 E-DVYPGDANGFVRHKSKTTPPASSGRGQAPLTOHYRYKALAPLAGNAITLNPVIEH 508  
QY 524 ----IEERHNSATQNLNSTLFQYNTDSELGRLKKQTECTGNGKTSYV---HKFTYTK 577  
DB 509 SETLTTHANPAHPLKEEKIYYLDAPKSSL-----RGRRYQEVVKKKLETTT 556  
QY 578 Q-----DDTLQOSSHSTTHDNFTIHRSOVRSRYGRLFSDPTKDIYVQNSYD 625  
DB 557 QYQFNSLIDPLGGHQVLETKTLFGYDGAQSTYQRRSLHGEELVELNENGVAITYQAYD 616  
QY 626 KLGRLTTLTNSGTPPANTLTYYELANLQDDNRPPVITTYDNGQLANEPDGAARHV 685  
DB 617 ALRRVTEERYSPTPEAKKRYDYOLCASDADLARAARV---TNAAGITTELEDGLRPT 673  
QY 686 SOCLKID--SDGDKFVYIHTQYDEQGRHHTSTYSYDVLTNGRQOTPDCKYHLSMSKSYDN 743  
DB 674 RESRDVLEARKPPGAFETILAIQIDAVGNRIQESVTWLO--AQO---YHLVTKRYDD 727  
QY 744 WQGIANTHMSYGVSEKITVDPI-----TLTATKQLOS-NSNNVQ----- 781  
DB 728 WGEQCCITIGPDVQGNHVLDPFGNADHQAIKYSWREGRLKSRPYQVNRNGMNVRSRQ 787  
QY 782 ----TGKEVTTYPSQOPIQITLPDEAGHLQSCHTLFRDQWDRKETTDAIGCTIYOYD 837  
DB 788 ARAISGTEETWMLFNKPVARKLDALGELIGERRYSYDGLGRTLTTERNHTTAESYD 847  
QY 838 NYNRVIQITLPDGIYVRKYAPFSTDTLT-----DIRVNGISLQQTFGDLSRLTQS 890  
DB 848 AWGRMLTTEQPNKTLIRTYAHSABDLPTRLVTPANVQLPARQIAQVFDGDLRLIGT 907  
QY 891 QDGGVVAITYASGNDQCPSTVITPDGQFIHYQOPELDAVLQVANSNETTOFSYNPVT 950  
DB 908 TTGRTRETYLPRDG-ESIPROINPAGETIEQDVNLQLTMEPISMTAPEEHSAPYAPVS 966  
QY 951 GALLKAVAEQSLPTIYPSGRKLKMNINMKKSYMTLRGLENGTDLTGTQKISR 1010  
DB 967 ARLLSDNRQCARFENKANKQLAEHMEW-KRQGWTSKSHSSTLQDLRNKTHEPYGED 1025  
QY 1011 T-----HGRVQLKDSIKTLYNYDDLNRHIGSQVTDLATGMLTTTVEFDGLNREIGR 1064  
DB 1026 TTHETDQAGRLVSTLQQLDAEFKYDDLGRILBTSHDPRASQALFKIEVDQDREYKR 1085  
QY 1065 KLCDSGHTLIDIQSMKLTQOLANRIVKNGVLQRTQOYSYSNRLNQYCKDAECPDT 1124  
DB 1086 TWKQGPQERTQETVMDKDLISRTLQVGVSVLVEKFGVSHARLNMVCTSPDQPRD 1145  
QY 1125 KYGHSYVTONFTYIYGNITPACTTTPADG--TEDHATKFKANPTPCOLTEVNHHPKMD 1183  
DB 1146 ALGRSIAMQVFNFAVNNIBLTVTSFTGPPAERATFTIME-RDPCQLRLTYVPPRTAP 1204  
QY 1184 NIRLKYKAGVINITNHNGBTENFTYDLGLT-----QNGQSVYGYDPLNRLYSOKTDT 1239  
DB 1205 NPEFSYDANG--NLTRDE-QARPIRYDSQRLRLGLANDSGAPDYGYDAGGLVSRPAG 1260  
QY 1236 KONSQERAYSAYGKHKSTANDASITLGYNGEADVDSVTHLGNQYRSYDPTLMEFHPDS 1355  
DB 1321 AGSTRAYRYTAVGERHADDPGLTGLTYNGEALDPSGMYLLGSGYRAVNPVLMRHSFDA 1380  
QY 1356 LSPFGAGGINPYSCLDPIRNSDPSGHL-----WQ-----AWT 1390  
DB 1381 LSPFGAGGLANNYGCGGNPIFRDPTGHSYIGYSGQSRSLADNLSYSIWRKALGALQWT 1440  
QY 1391 GIGMGI-----AGLLTLATGMA-----IAAGGIA-----AALASTS--- 1424  
DB 1441 GIGIGILFAAVVAVAVVVTGVAAPALIAAMAAGGIIISGAGALISAGVGLISVSLAV 1500  
QY 1425 ----TTALAPALSVTDISITVSGALEDASPKASSILGWSMGAGLAEASAIKSGTK 1480



Db 1501 GIKLAAVATATLSVTG--TALQTEALISGHEKKNIGTILNYSAILGLAVGMOQLIAK 1558  
 QY 1481 ---LATHGAPAD--GEMNLKSTSSSRKMKVTSLDREIV 1519  
 Db 1559 IPNLWAPAGSYTTTITDPLINSKTLSELK-----MAKQII 1596

RESULT 11  
 Q45948 COXBU  
 ID Q45948 COXBU PRELIMINARY; PRT; 709 AA.  
 AC Q45948;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE 709.  
 OS Coxiella burnetii.  
 OC Plasmid OPH1.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;  
 OC Coxiellaceae; Coxiella.  
 NCBI\_TaxID=777;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Thiele D., Willems H., Haas M., Kraus H.;  
 RT "Analysis of the entire nucleotide sequence of the cryptic plasmid  
 OPH1 from Coxiella burnetii."  
 RL EMBL, X75356, CA53129.1; -; Genomic\_DNA.  
 DR PIR, S38241; S38241.  
 DR InterPro; IPR006530; YD.  
 DR Pfam; PF05593; RNS repeat; 8.  
 DR TIGRFam; TIGR01643; YD\_repeat\_2x; 4.  
 KM Plasmid.  
 SQ SEQUENCE 709 AA; 79416 MW; A37789A2A72CCF86 CRC64;

Query Match 13.7%; Score 1214; DB 2; Length 709;  
 Best local similarity 40.8%; Pred. No. 2.2e-57;  
 Matches 287; Conservative 101; Mismatches 280; Indels 36; Gaps 16;

QY 554 KQTECTGKNGKTYSVNHKTYTKQDDTLQOSSHITTHDNFTIRSGVRSRYGRFSDT 613  
 Db 3 RRAVLTSSKSGKTYQONQTTFAISSQAMHLQKIDFTGDDGKITSISREGRYSGRLLST 62  
 QY 614 DTDIDIVQMSYDKLGRLLTTRTLN--SGTPYANTLTYYELNLLQDDNRPPIVITTVNGN 672  
 Db 63 DELGANTQYEDDELGRLLTQTVANASTTYASTTYSISLELDAGKATKATTVTDKGN 122  
 QY 673 QLNREPDGAGRNVSCDKSDG----DGRFYTHITQYDEQGHHTSYSDVLTNGRQ 727  
 Db 123 QLRITYDGLGRNLKQERLDDAASQKGTWYITHQOYDALGRESKTIQDVL-----R 177  
 QY 728 TDPDKHA---LMSK--SYDNWQIANTHMSYVSEKITVDPITLTATQLOQSNMNVQ 781  
 Db 178 LDBEVGAGSVLSTSKMHWDSWQNHLYFSDDYQRSYVDPIRAT--LQPESSQOK 235  
 QY 782 TGRKVTYTPSOQPIQITLTFDEAGHLQSCHTLTFRDGDRVKEKEDALQGCCTIYDNYNR 841  
 Db 236 LGGQLTFYNLAGRPKIKTYQYDSQTEGSAHYEDGQLRKEDDELQTLTLEYDIFGR 295  
 QY 842 VIQITLPDGTIVNRKVAFFSTDLITDIRNGISLGGQTFDGLSRLTQSDGGRVMAVY 901  
 Db 296 VTQITLPENTIIQGSYVAPHSATSLTGISVNFMSGQTFDSLRLTETTSGGRTAFSY 355  
 QY 902 SAANDQPSVITPDGFIHQYQPELDVAVQVANSIEITQPSYNAVNTGALKAV--AEG 960  
 Db 356 ENNS--SVPAAVNAPVTSVSEYELKEIGNAVKISABEIIQVWDYDLTGAMTSATQAG 414  
 QY 961 QSLTFPIYPSGRKME--NIND--MKMSYLMTLRGLNGEYTDLTGIIQISRTTHGRVT 1016  
 Db 415 MIQGMITYPSGLKNERSMPSDGAOKSTATTYSIAGAPQSTIVFGTQYVYDDEHGRKI 474  
 QY 1017 QIYDSIKITLANTDLNRHIGQVTDLATGHMLTTVEFDGLNREIGRKLCDSSGHTLDT 1076  
 Db 475 GIEBNDIKVSLDYDAFGRFTKQATDKKGAVALSTLTLYDLNREIGRISASQSVLVI 534

QY 1077 QQSWLKTQOLANRIVKLVGLQRTQYSDSRRLNNGYKCDGACPTDKYGHSTVQNF 1136  
 Db 535 EQYQRNHLKERTQGRITLREKEMFAYDSRNRLIEYTCNGEARPODPYGAHROTF 594  
 QY 1137 YDIYGNITACHTTFADTETHAFKFPANPDPCQLTEVHHHTPPMPNIRLYKDKARVI 1196  
 Db 595 YDALGNNTKQOTPESSG--RNTATYIY--SALDPQLLVNNDHSPKPEITLEYKAGRM 652

QY 1197 NITDNHNTENFTYDYLGRLO--NG---QGSVGYDPLANRIVSQ 1235  
 Db 653 R--DEAGRT--LRVDALGRLOVNGAGKGGQYAVDVIALTVISQ 692

RESULT 12  
 Q88LP2\_PSEPK  
 ID Q88LP2\_PSEPK PRELIMINARY; PRT; 1290 AA.  
 AC Q88LP2;  
 DT 01-JUN-2003 (TREMBlrel. 24, Created)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE Hypothetical protein.  
 GN OrderedLocustNames=PP1887;  
 OS Pseudomonas putida (strain KT2440).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; Pseudomonas.  
 NCBI\_TaxID=160468;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RA MEDLINE=22423060; PubMed=12534463;  
 RX DOI=10.1046/j.1462-2920.2002.00366.x;  
 RA Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,  
 RA Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,  
 RA Brinkac L.M., Beanan M.J., DeBoy R.T., Daugherty S.C., Kolonay J.F.,  
 RA Madupu R., Nelson W.C., White O., Peterson J.D., Khouri H.M.,  
 RA Hance I., Chris Lee P., Holtzapfe B.K., Scanlan D., Tran K.,  
 RA Moazzaz A., Ultebeck T.R., Rizzo M., Lee K., Koback D., Westl D.,  
 RA Medler H., Lauber J., Stepanovic D., Hobeisel J., Straetz M., Helm S.,  
 RA Klewitz C., Bisen J.A., Timmis K.N., Duesterhoeft A., Thiemler B.,  
 RA Fraser C.M.;  
 RT "Complete genome sequence and comparative analysis of the  
 metabolically versatile Pseudomonas putida KT2440."  
 RL Environ. Microbiol. 4:799-808 (2002).  
 DR EMBL, AE016781; AAM67506.1; -; Genomic\_DNA.  
 DR TIGR; PP1887; -;  
 DR InterPro; IPR006530; YD.  
 DR Pfam; PF05593; RNS repeat; 5.  
 DR TIGRFam; TIGR01643; YD\_repeat\_2x; 4.  
 KM Complete proteome; Hypothetical protein.  
 SQ SEQUENCE 1290 AA; 146578 MW; 7081A3F10B287F92 CRC64;

Query Match 11.7%; Score 1037; DB 2; Length 1290;  
 Best local similarity 27.7%; Pred. No. 2.5e-47;  
 Matches 375; Conservative 196; Mismatches 578; Indels 204; Gaps 52;

QY 14 MSDNNEFTQANNFTSAGVNDPRGLNVIQITLGHIVNGMLGPTLPLTSLPLNKT 73  
 Db 2 MSTSVSHASNAFSSSTIESGVDPRGQYVSRILPELQNDLQGFELALFSPPLNGE 61  
 QY 74 DIFGIGFNGSLGVYRKNSLTLSTGENYKVI--ETDKYKLOQKLDNLRPEKDLKEN- 131  
 Db 62 DSGFGKQWN--LDLTQVRKHIYVLTSSGETYKIKIGKSSVTRLEMKRKLQRF--DYEDP 117  
 QY 132 -----CYRIHKSGDIEVLTFGNNAFPLKVPKILNAPAGNAIYIDW--NFEATQRLN 183  
 Db 118 PCPGAARFVHERSGVLEVLVWGSGEQEVALLPVELHSLPGRHLHSYLPFGQSHRLS 177  
 QY 184 RIYDDLGDHDIPLNLEYQGLIKITLTLFPGQGEGRTEIRF--LNROANSIHNFSLGN 240  
 Db 178 EVQDSQD---VLRIQRSNDSRVELLCPYSGSDGQPLARVAMTLSESQNRVSEIILPT 233  
 QY 241 ENPLTWSFGYTPIGKNGIILQ--WITSMTAPGGLKETVTVSNNGHHPQSANLPLV-LPY 298

Db 234 ANQARAFETY-----EDVGLVCVRECTTPYGVVERVY---QDAGHKFPSSAARDKXLP 286  
Qy 299 VTLMKVPGAGQPAIQAEISYT--SHNYVGGSGSNCIMNNK-LDNIYGLMTEYNTGSTESRR 356  
Db 287 VTRHEIDPRGQAKVVRVEYPTGTHNFIQGGSSISMSDDGIDNIYKVEBYTYKSI01-- 344  
Qy 357 YKDEGHDOIVR-1ERTYNNYHLLTSECKOONGYIQOTTEPAVYALISHANPDSOPSOPL 415  
Db 345 ---OEVAROSVRTTTRTFNPHLITEQATQAGDQLQAFRYADNAG-NESQSPRYQLP 400  
Qy 416 KTKETWRSADN--SYRSEITETTFDESGNPLTKVYIKDKTKQKISBSTMERYYPAGEV 473  
Db 401 HDETFQMSLISNPERQREBKRVTKYDGHGVLFTLLPNQVLETNV-----WYSABEEG 453  
Qy 474 DNECPREYGTFRPK-KIIQ-----TPYDSEFKODPEKFIORY 511  
Db 454 DE-----HGFVRNLTCTVQPAATGHAATAITGAFOYRALTPPGS-----YLQOPW 500  
Qy 512 SLIGSOSHVTLKIEBRYSATQULN--STLPQ-VNTQKSELGRLLKQTECTKG-ENGKTY 567  
Db 501 RLUESET-----LSESGSAPETFEKISKLYQESASERFSYGRVQOTVSYPGVDGSPF 555  
Qy 568 SVNKEFTYTKODD--TLQOSHSTITHDNFTIHSQVRSRYTGRLLFSDTKDIYQMSYD 625  
Db 556 DTLTHYGSLPDBERALTQVEKLVGVDOESKTIILRHALDTGEBPLMLDNGVEIRYYD 615  
Qy 626 KLGHLTRTLNSGTPYANTLTLYDELNLODNRPPVITTTDVGNGQLRNEPFGAGRIV 685  
Db 616 ALRVTREVAAPGRFPEARNYKFLCAYDNEQAQMAVYQVETHTL--LDGLSRPI 672  
Qy 686 SOCLKSDG---DGEFYTIHTQYDEQGRHNTSYSDYLTNGRQOTDPDKVHLSMSKSYD 742  
Db 673 FEEBADDSATYAALAPRIYARDELQOLVEBTEIDMLSGG-----LLELTSHSYD 725  
Qy 743 NMGG---IANTHWSYGVSEKITVDPITLT-----ATKQLOSNNNVQTEKEVTTYPSQ 793  
Db 726 DMGQYAVANLPDGTQVEE---IDQVASTDGPVHARAMEVE---HSRVSIGITETWMLFE 779  
Qy 794 QPIQITLFDENGLQ-SCHTLTRDGMRVKRETAIG---QCTIYQVNDNRVQIITLPD 849  
Db 780 KPVRIERPALDGSSTISLQVNDYDGLGLSREAGSGAGQGVNVEYRDAFRAVEERLAD 839  
Qy 850 G-TIVNRKAPPSYDITLDIRV-----GISLQCFPEASRLTOSODGGRWAAVYTYSA 903  
Db 840 RYNTVYRTYAGHSNDLPVSIKQNTBSAVALGBOYVDSGLERTVAITGSRBQTFEEDP 899  
Qy 904 GNDQCPSTVITPDGOFIHYQYQPEL-DDAVLQVANSNEITQOFSYNPVTGALLKAVAGQS 962  
Db 900 GERQ-FHVVKAPDGRITIEYQYRPALGEBPVLRLSGK-EAKYEDLKNAARLTHCEEPDD 957  
Qy 963 LTPY-----YSGRLKMN-----INDMKMSYMLTGLGNGCYTDLTGTIQKISDTH 1012  
Db 958 ENSGTYLDRSHPLSNGEVRKRESRYVDGEAFSMTYDYSFRSLRAVVDLGGTQLYDPFDV 1017  
Qy 1013 GRVTOIKDS-----IKTTLNVDLNRHIGSQVTLATGATGMLTTT 1052  
Db 1018 GRLEKTLHAPEPKTRRYRLQARARQULLLESTFGYRQGRMASITTTDASTGHALATL 1077  
Qy 1053 VEPGLNREIGRKLCDSSGHTLIDIOQSWLKTQOLANRIYK-----LNGVLOTREOYS 1104  
Db 1078 LEYBEFDEILRTF-DFGDTYQTLAQDYDEPDKLSRIKERPKXSDESQATLLRHEYYQ 1136  
Qy 1105 YDSNRLNQYKCDGAECPYDKYGSIVYONFTYDIYGNITACHTFAGCTEDHATFKKAN 1164  
Db 1137 YDRGRRLQIYTCDEAPVDSGQITARQIFGFDLNDILSVITYRPGSQRILTVEKEN 1196  
Qy 1165 PTDPCQ-----LTVHHTHPDMPD-----NIRLKXKAGRYINITDNHGN 1204  
Db 1197 -SDPQMSRIIPRDLLETYVDVHDLBELFKI VAKDPQLADLHNDGNGN--ISDBQGR 1253  
Qy 1205 TENFTYDTLGR-----ONGGSGYGYDPLNRL 1232  
Db 1254 V-LTYDGLNRLLRVETPDGERCRYNDPENIL 1284

RESULT 13  
Q883W6\_PSESMS  
ID Q883W6\_PSESMS PRELIMINARY; PRT; 982 AA.  
AC Q883W6;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE YD repeat protein.  
GN OrderedLocustNames=PSPT02231;  
OS Pseudomonas syringae (pv. tomato).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=323;  
RN [1]  
RP NOCLUSTOTIE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RC STRAIN=DC3000;  
RX MEDLINE=22834015; PubMed=12928499; DOI=10.1073/pnas.1731982100;  
RA Buell C.R., Joaridar V., Lindeberg M., Selengut J., Paulsen I.T.,  
RA Gwyn M.L., Dodson R.J., Deboy R.T., Durkin A.S., Kolonay J.F.,  
RA Madupu R., Daugherty S.C., Zafar N., Zhai L., Liu J., Yuan Q.,  
RA Nelson W.C., Davidson T.M., Zafar N., Zhai L., Liu J., Yuan Q.,  
RA Khoui H.M., Fedorova N.B., Tran B., Russell D., Betsy K.J.,  
RA Uetexback T.R., Van Aken S.B., Feldblyum T.V., D'Ascenzo M.,  
RA Dang W.-L., Ramos A.R., Alfano J.R., Carlinhour S., Chatterjee A.K.,  
RA Delaney T.P., Lazarowitz S.G., Martin G.B., Schneider D.J., Tang X.,  
RA Bender C.L., White O., Fraser C.M., Collier A.;  
RT "The complete genome sequence of the Arabidopsis and tomato pathogen  
Pseudomonas syringae pv. tomato DC3000."  
RL Proc. Natl. Acad. Sci. U.S.A. 100:10181-10186(2003).  
DR EMBL: AB016853; AAO55747.1; -; Genomic\_DNA.  
DR TIGR: PSP02231; -;  
DR InterPro: IPR006530; YD.  
DR Pfam: PF05593; RNS\_repeat; 6.  
DR TIGRFAMs: TIGRFAM01643; YD\_repeat\_2x; 2.  
DR Complete proteome.  
SQ SEQUENCE 982 AA; 106800 MW; 00E7CA6B93E19B78 CRC64;  
  
Query Match 10.5%; Score 932.5; DB 2; Length 982;  
Best Local Similarity 29.7%; Pred. No. 7.9e-42;  
Matches 285; Conservative 142; Mismatches 357; Indels 175; Gaps 33;  
  
Qy 642 ANTLTYDEELNLODNRPPVITTTDVGNGQLRNEPFGAGRVSQCLKSDGGEFYTI 701  
Db 31 SSTIGRYD-----DMNGR--CCTTTDDN-VQYVEYSDPIGSDVHK-----GPIQKT 74  
Qy 702 HTQCYDEQGRHNTSYSDYLTNGRQOTDPDKVHLSMSKYDNMGQIANHWSYGVSEKIT 761  
Db 75 WKQSGDPEGR-----ISGSEF-----MLNT-----FGKPRIR 103  
Qy 762 VDPITLTATQLOSNNNVQTEKEVTTYPSQPIQITLFDENGLQSCHTLTRDGMDRV 821  
Db 104 ---TLTAGTGRSHRTSMRSRRLTY--TEQEL-----SRQTFYDGLGRC 144  
Qy 822 RKETDAIGCCTIYQYDYNVNVITQITLPDGTIVNRKVAFPSTDLITDIRV---NGIS--- 875  
Db 145 TEGEDALQOSTFIEYDWSMWSSTLADGSVINRSYAPQSSSELATLLELVHNGTTRTV 204  
Qy 876 LGOQTFDELRLTQSGODGRVWATYTSAGNDQCPSTVITPDGOFIHYQYQPELDAVLQV 935  
Db 205 AGTQKFDGLRVRVQTKYGDVEQGFNYDAGEMQ-PRSRITAGLDINIFTYATLTDQIPSS 263  
Qy 936 ASNEITQOFSYNPVTGALLKAVAGSGLTPIYSSGLKXENINIMKKMSGLW-----T 989  
Db 264 TAPDETAKFYDKTSABRLIEATNPQGRTRYRYDVHNDLTGFTWNL--LQAMETRHQS 321  
Qy 990 LRGLNGYVTDL-----TGTIQKISRDTGRVTOIKDSIKTTLNVDLNRHIGSQVTDIA 1044  
Db 322 ILGRPIRTDLKKEAAGAEIRYVDYTLGRIRFINGNRLRTIIDYDVLGQCYKXATDELQ 381  
Qy 1045 TGRMLTTVEPFDGLNREIGRKLCDSSGHTLIDIOQSWLKTQOLANRIYKLVNGLORTREYS 1104

```

Db 382 AGTGVIIIDMEYDDOGGEIILRTQTASNOALTLTQTMAVDGLKTRDLQAGSPILHETFS 441
Qy 1105 YDSRRLNOKKCGAEPTKXGHSIVTQNFYDIYGNITACHTTPADGDEHATFEKFN 1164
Db 442 YDRGRRLTVNVLGSSLPDELQREMTQIFSPDELINILTCQRFDTGTSERAAFKXS 501
Qy 1165 P-----TDPQOLTEVHHHTHDPMDPNIRLKDKAGRVINITDNGNTENFTYDILGRL--- 1216
Db 502 PGDDKHDKDRQQLSIATVPKRPDPPTFSIDANGN--QLKDEHGN--SLHYDSQSLQLQV 557
Qy 1217 -QNGQG--SVYGYDPLNRLVSOKTDLTDCBL--YYRETLMLVNEVRNGEMIRLRTGETI 1271
Db 558 AETGAPAIQGYRYDGHQVLVATR-DGNESHLIRYEGHQSLSTVQDQRTOYLHGEQPL 616
Qy 1272 AQC---RASKVLLTGTDSSQSVILTSKQNLSCQAYAGYKHKSTANDASILGNGRAD 1328
Db 617 GQGIIVDAEQTLLLTLDANQSVMEGFQOQLRKAVYSAYGBRHSBALSTAGNGEVR 676
Qy 1329 PVSGVTHLNGYRSYDPTLRFPTPDLSPFGAGGIPYSGYCLGDPINRSDPSGH----- 1383
Db 677 AANGWILNGYRAYNPLNRFHSPDLSPFAGGVNPTTYCLGNPLALMDPTGHDASQ 736
Qy 1384 -----LSWQ-----AMTIGMGIA-----GLLTITAGMAIAAGGI----- 1416
Db 737 TGRLRPDEGALPMQGGGDIMGVVGVIGVFTVLGVAATITATIGATPTVGPVTLGI 796
Qy 1417 -----AAIASSTITLALG-----ALSVSDITSYSGALEDSPPASSILGVS----- 1462
Db 797 SMTASAAAVSTVGTALIVGTALTAATANTVAIVNN-----DQTAGEGVLGIAAV 851
Qy 1463 -----MGGAAGLBSAIIKGGTKLA-----THLGAFAIDENALLKSTSESSR 1505
Db 852 PVGLVFGAGAVARAAVAAAATAAANAAGTIGVRSVRIG-LAAAGARRTISMAASAR 909

```

## RESULT 14

Q87VG6\_PESM

ID Q87VG6\_PESM PRELIMINARY; PRT; 1669 AA.

```

AC 01-JUN-2003 (TREMblrel. 24, Created)
DT 01-JUN-2003 (TREMblrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
DE YD repeat protein.
GN Ordered locus names: BSP04970;
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OX NCBI_TaxID=323;
[1]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RP STRAIN=DC3000;
RC MEDLINE=22834015; PubMed=12928499; DOI=10.1073/pnas.1731982100;
RA Buell C.R., Joardar V., Lindberg M., Selengut J., Paulsen I.T.,
RA Gwinn M.L., Dodson R.J., DeBoy R.T., Durkin A.S., Kolonay J.F.,
RA Madupu R., Daugherty S.C., Brinkac L.M., Beaman M.J., Haft D.H.,
RA Nelson W.C., Davidson T.M., Zafer N., Zhou L., Liu J., Yuan Q.,
RA Khouri H.M., Fedorova N.B., Tran B., Russell D., Berry K.J.,
RA Utecherack T.R., Van Aken S.E., Feldblyum T.V., D'Ascenzo M.,
RA Delaney T.P., Lazarowitz S.G., Martin G.B., Schneider D.J., Tang X.,
RA Bender C.L., White O., Fraser C.M., Collier A.;
RT "The complete genome sequence of the Arabidopsis and tomato pathogen
Pseudomonas syringae pv. tomato DC3000."
Proc. Natl. Acad. Sci. U.S.A. 100:10181-10186(2003).
DR EMBL; AEO16853; AAC08398.1; -, Genomic DNA.
DR TIGR; PSEP04970; -.
DR InterPro; IPR006530; YD.
DR Pfam; PF05593; RNS_repeat; 5.
DR TIGRfam; TIGR01643; YD_repeat_2x; 2.
KW Complete proteome.
SQ
SEQUENCE 1669 AA; 184268 MW; BC668B6C1E3FB7EE CRC64;

```

Query Match 10.4%; Score 920.5; DB 2; Length 1669;

Best Local Similarity 25.9%; Pred. No. 8,1e-41; Matches 440; Conservative 212; Mismatches 662; Indels 387; Gaps 84;

```

Qy 152 AFDLKVPKLLNDPAGHAIYIDNPEATQPRLNRIYDDLGDHIDPLNLEYQGL---IKTI 208
Db 72 AFGQLLPSEVTSFQGGTLSTIEMSRNQQFLKKI---SSDANI--LSTYSABSDSTSTV 127
Qy 209 -LTLFPOQGEYGTBLRFLNRQLNSHNSLGNENPLTWSEFGYTPGKNIILQGMITSMT 267
Db 128 EISVWPETDAVYVRLSELSSVALTGLRIEKNLKVQOKLKYGC---ADPTLRVLRRIE 184
Qy 268 APGLKETVYVSNNOQHHPQGANLVLPEYTLMKQVPAGAPAIQAEYSYSHNYVG 327
Db 185 EBDGSIELVYTRRG--GMPEPYRQ--PLPCVTLHSLFPBACSNITDHFYGGTYNLGF 241
Qy 328 GSNGINMKLNDLYGLMTEVYNGSTESRRYKDEGH-----DOIYRIERTNNYL 378
Db 242 SEPP--DAHQNRLY-----YERLELRLELVDEGYVQILRQNPDIIVSTRHAFNRNHL 292
Qy 379 LTSECKQGNQYIQTETAYYAIIGHNFDSPQ--POLPK--TKETWRBADSYRSEIT 434
Db 293 QVRE-----DLQVQFAEKVISMEPANASPGKVFGLPYKTIYDYDLSHPNTERITTV 346
Qy 435 EITFDESGNPLTKYIKDKTKTIISPSTHWEYYPAG---EVD-----NC 476
Db 347 QTLANNIGQLTKSI-----AVDGVTEMLYTPDQGGGLDLSIAERKSLDVLTLTC 400
Qy 477 PPEPYGTFRKKIIGTPYDSEKDP---EKFIQRY-----SLIGS 516
Db 401 PKVSEGY-----MPVVEYVHDPAYPKQOITAYAYQARENPNQRSVLVPSYVV 452
Qy 517 QSHVTLKIERHRSATQNLNSTLFOYNTDSEBELRLKQRECKG-----ENCKTY--S 568
Db 453 LTGVTL-----DTTMLPSLM-----BGRNALIEQVIRSLIPDVATEVTWAKES 499
Qy 569 VVHKFTYTKQDDTLQOSSHSTTHDNFTIHR-----SQVSRKYRFLPSDTDK 616
Db 500 VVQNSMLGMQRNLT--TTSMLYDNDPSVGTVAABQKLISSRIFSRLSGRPLSER--RD 556
Qy 617 DIVQMSYDKLGRLLR--TLNSGTPYANTL--TYDEI--NNLQDNRPPVITTDVN 670
Db 557 GLRFRVYHDSIGRLVQERGTBAGMKADAVETTVYSITAEGLQ-----VTVEAE 608
Qy 671 GNLGRNEFDGAGHYVQCLKSD--GDKFTYTHQYDEGRHHTSYSDYLNNGRQTD 729
Db 609 -QQVRLYLQRPVWVAIKRLLPISAFVIRIEVDGPDATNQTL--DYLPGGLRRTK 666
Qy 730 PDKVHLMSKSYDNMQIANTHWSYG--VSEKITVDPITLTATKQLOSNNNYQTKREV 787
Db 667 DAREPAVNDASKLAW--MADYTRBDAGILINEQVIGADSGAQLRQLSGRLDLSHTALLE 725
Qy 788 TYTPS-----QQPIQITLFEAGHLQSCHTLTRDGMVRKRETDIAIGCTTYQYDNYN 840
Db 726 TLAPSTARDASTRTIERT--FDDERLLIKIRT-----SNTSEHC--IEYDELE 770
Qy 841 RVQITLPDQSTVNRKKAAPSTLTLLTDIRVNGISLGGQTFDGLSRLTSGQDGRWAYT 900
Db 771 RAAVAILAPDTRTERKTHOLS--DYITQLNVGSTVIGTQMTAAAGQTVGE---LTYE 824
Qy 901 YSANGQCPSTVTPPGQFIHYQYOPELDAVYQVANSNETTQO-----SSYNPVT- 950
Db 825 FPGGS---ASTVRPDKTILLESASVADGHTATLSINKKHTQAVIGSPNVLJYTVDPVS 881
Qy 951 -----GALLKAAVAGGSLPIYV--PSGRKLMENINDMKKMSYIWLTLGLNGYTDLTGT 1003
Db 882 PSAEAMSLTSSPQSIGLTSITQTSPPGRSQAMTRSLKG-----RLLTN--TAVDGR 932
Qy 1004 IOKISPDTHGRVYQIOWSSIKTILANDDLNRHGS-----QVTLANQMLTTIVRPGIN 1059
Db 933 QMKVFYDYLDVRVYV---ILGELHYHLYMSAFGBLPQRTVAVQASERLDVRFYWDVAFG 988
Qy 1060 REIGRKLCSSGHTLDIQQSWLKTQOLANRIVKLVGLQRTQEQSYDSRRLNQYKC--D 1117

```

```

Db      989 QE1AREYTLNKKPLLATNSYLANQVSKTLTREBVLQRTBEGSYDARDRLNSYECTTD 1048
Qy      1118 GACPEPTDKGHSTVTONFTYDIQNTIACHTTPADGREDIATKFK--ANPTBQOL-TE 1173
Db      1049 VAWMPQDQAKSLSKQSYGYDELHNLSECSYADGSTCIQTYTYDTPVKNPTRLRLSKYTE 1108
Qy      1174 VHH-THPDMEDNRLKTKDKAGRVINITDNHNGTENFTYDTLGRLL--QNGQGSV--YGY 1226
Db      1109 LRSGQSTTSQOTATLAYDANGN--QTTDESGRT--LAYTPGGQALASKYKNDGKLLTRYXY 1164
Qy      1227 DPLNRLVSQKTDLT--DCELYREFTMLVNEV--RNGEMIRLRLT-----GETII 1271
Db      1165 DAFRLISQYIGATKHTCCELLYDGTQLTGEAMFPDARREFRILFSBDMVQOTTCIGETV- 1223
Qy      1272 AQRASAKVLLTGDSQOSVILTSQK-----NLQGEYSAVGHKSTANDAST-IGVNGE 1325
Db      1224 ---RSDVVL---TDPGSGVGFSAHDGTAGVKLHPLGTYPYG--ESTNLSGCRLGENSE 1275
Qy      1326 RADPVSGVTHLNGYRSYDPTLMEFHTPDSLSPGAGINPVSYCLGDPINRSDPSGHL- 1384
Db      1276 RIDPVLGMYHLNGYRTYSPAQRHMLQPDSSPFGAGINNTAYCAGDPVNLDPBSGHVM 1335
Qy      1385 --SWQA-----WTGI-----GMGIAGLLLTATGGMALAAAGI 1416
Db      1336 ISRKGAGNMTSDLTKALQETSPPQLGHPMRLAVNASAVAGVLMVPLTGSSSLGFAAGV 1395
Qy      1417 AAATASTTRALAGALSVTSIDTISYGALEDSPK-----ASSILGVSMGMAG 1469
Db      1396 LA-----TTLAVASAGL-----DIASVY--LEDVNPBLARKLGTAAALGFIISNAPFRAG 1443
Qy      1470 L-----AESAIKGTGLATHLGNF--ABDGENALLKTSSESRK 1507
Db      1444 LRLGRLRLRTSSRTGRLISVESITFKGPKLLTGITIKLKPQ-----MSYASNVTR 1499
Qy      1508 WGVTRSLDREIVRNEEGOVIKDHSRGYTDNFMKGEOAILVHGXDG---FLYHTEGK 1563
Db      1500 W-----APKLDKF--EKLBNALFMTVDKQGGKRVTFMAH--GVK 1534
Qy      1564 -----HNGKGYT--RHTRPEQLVDLXKNNTYDLVQGGDKPVLHLSYGS 1607
Db      1535 PDDVGDAAAMAIIEYSGSPSGSKSFSAYSPFDLKSCK-VDLNK--YEVKRLIMCHSAD 1591
Qy      1608 SGA---ADKMAKYINRPVIAY 1625
Db      1592 GGEKSPATFSKLTNKPVKGY 1612

RESULT 15
045905 COXBU
ID 045905 COXBU PRELIMINARY; PRT; 505 AA.
AC 045905_
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Orf 505 protein.
GN Name-orf 505;
OS Coxiella burnetii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
OC Coxiellaceae; Coxiella.
OX NCBI_TaxID=777;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=97294470; PubMed=9150226;
RA Willem H., Rutter M., Jager C., Thiele D.;
RT "Plasmid-homologous sequences in the chromosome of plasmidless
RT Coxiella burnetii Scurry Q217."
RL J. Bacteriol. 179:3293-3297(1997).
DR EMBL; X93204; XA63682.1; -; Genomic DNA.
SQ SEQUENCE 505 AA; 57841 MW; 8E60E3C307CABD0B CRC64;

```

Query Match 10.4%; Score 919.5; DB 2; Length 505;  
 Best Local Similarity 41.1%; Pred. No. 1.4e-41;  
 Matches 217; Conservative 75; Mismatches 195; Indels 41; Gaps 16;

```

Qy      18 NEF-FTQANNFTSAVSGVDPRTGLXNIQITLGHIVGNGLAGTLLPLTSLYSPLNKTDI- 75
Db      2 NELPYQAINFTSAVGGVDPRGTGLFTVMNVAELGNDNLGDFLPTLNTSLSLSNIC 61
Qy      76 GFGIGNFGLSVYDRKNSLISLSTGENYKVIETDKTVKLOQKLDLRFPEKDLKENCYRI 135
Db      62 GFGIGGSVGISYDKNNKLLISSLSSERYKTEDNDSVYVQKKINNFKEK--IKNGYII 119
Qy      136 IHSGDIEVLTLGNNAAPDLKVPKLLNPAGNAIYIDWNEAQPRLNRIYDDLDDGHDIP 195
Db      120 KYNKGTEYLKYGDNLP--LPQKIFSLGMPDKLSWENRGQYVNLTKLEDAD-----V 172
Qy      196 LNLLEYQGLIKTILTLFPQKKEGYRTLEPLNQLNSIHNFSLGNENPLTWSFGYPIGK 255
Db      173 LCKIDYQFSMAKITTPWPKTESYTFQLDPVNELYLWVNTKSLRSR--LWMSFNVDVGA 230
Qy      256 NGILGQWITSMTAPGGIKETVNTSNNGHHFPOSANLPVLPYVTLMKQVPAGQPAIOA 315
Db      231 GNFT---LTQVKSPTGLTETVNYQAGVM--RFPDESCKPALPSVYVNYRQSPGMQPDIVK 285
Qy      316 EYGTSHNYVG--GSGNGIMNKLNDLYG--LMTETVNGSTESRRYKXKQEGHDQVRIERTY 373
Db      286 EYETASNYLYGASLSGKANEDBDNTYVMDQYTVSSTEKLIYDNRE---LVSISRIY 341
Qy      374 NNYHLTSECKQONGYIQTETAYVAILIGHNFDSPSPQFLPKTKETWR--SADNSYRSE 432
Db      342 NSYTLILSETTRQNSCEVIVETIYAKPGLSPKQPKQPLPKEBKTKWENSGNQCRSE 401
Qy      433 ITETTEDSGNPLTYVYIKDKTKOKIISPTHWBYYPAGVDN---CPPEPYGFTFRVK 488
Db      402 ITTTFDPEGNLLTKIEPD-----GTYKTEYIYYDSKGETDGIYVCPPEPNGFVAFVK 454
Qy      489 KIQTPEYDSEFKDDPEKFIQRYRS---LIGSQSHYTLKIEBHYSAT 532
Db      455 TQIVTPANSEFY-APVQQTYYAAYQYPCIASLSSLSIAVLONKGFAYT 501

```

Search completed: January 30, 2006, 09:49:02  
 Job time : 141.373 secs